DIALOGUE AND THE MACHINE: AN INTERACTIONAL PERSPECTIVE ON COMPUTER DIALOGUE MODELS, MEDIATION AND ARTIFACTS

by

Paul Bruce McIlvenny

Thesis presented for the Degree of
Doctor of Philosophy of the University of Edinburgh
in the Faculty of Science
1990

ABSTRACT

The topic of this thesis is the notion of dialogue and how machines have not only influenced the development of our understanding of this fundamental human social activity but also the possibilities for engaging in mediated dialogue. In particular, the concern is with its adoption and distortion from a computational point of view. An interactional perspective is developed that provides insight into the problems and limitations of computer dialogue models, motivates the investigation of the achievement of dialogue mediated 'through' machines, and informs the conception and design of computer systems (or artifacts) that support the metaphor of dialogue 'with' machines.

To motivate a reconstruction of the notion of dialogue and a different understanding of the status of machines in terms of action, a critical analysis of computer models of dialogue, concerning theory, data and implementation, is given. In general, computer models lack a consideration of interaction as a constitutive domain, assume the interchange model of dialogue, promote a sanitised view of data, and are a poor foundation for the design of machines that are to engage in dialogue-like behaviour with a user. An alternative interactional perspective is derived from hermeneutics and ethnomethodology in which it is argued that the machine is an intelligible - not intelligent - artifact, and communicative activity is circumstantial, situated and interactively constituted. Instead of reifying dialogue as the repeated exchange of discrete messages between isolated cognitive processors (the interchange model), dialogue is understood here to be the collection of practices in which parties are mutually engaged in coordinating communicative actions and achieving shared understanding out of the materials at hand. The empirical methodology of the thesis comes from conversation analysis and forms the basis for the investigation of the achievement of dialogue 'through' machines.

A detailed audio-visual study of a particular computer-mediated communication modality is presented. Parties engaged in cooperatively constructing mutual orientation in dialogue (in a virtual dialogue space) were recorded and features of their conduct were rendered for analysis with the aid of a notation system specially developed for this study. The findings are that the computer-mediated dialogue activity is a skilled, interactive accomplishment in which dialogic presence, monitoring and participation are contingently created and maintained. An emergent transformation of the dialogue activity demonstrates the situated work of constructing participation, a process that is shaped by the dynamics of that activity. A brief study of copresent collaboration documents two further features: the embodiment of actions and their complementarity. The consequences of the interactional perspective and the empirical study for computer models and dialogue 'with' machines are discussed. Suggestions are also made about an alternative use of computer modelling for dialogue 'between' machines, and about the future of dialogue mediation and artifacts.

Declaration

I declare that this thesis has been composed by myself and also that the research reported has been conducted by myself unless otherwise stated.

Oulu, Finland, 17th September 1990

Paul McIlvenny

ACKNOWLEDGEMENTS

Reconstruction of a rosy dialogic past...

I am indebted to my final supervisor, Jim Miller, who patiently gave support and helpful criticism for the latter and most focused half of my postgraduate studies. The first two supervisors were Peter Jackson and Mark Steedman who both tolerantly accepted the non-traditional and difficult course of the early stages of the thesis. I am also grateful to Paul Drew for his generosity in spending time to explain some of the workings of conversation analysis to me and to pass on many manuscripts about the field. I am also thankful to those who read all or part of the thesis as it neared completion: Jim Miller, Pirkko Raudaskoski, Jim Haines, Ian Morris-Wilson and Ivan Leudar. Their comments on style and content have improved the final version immeasurably. Of course all remaining errors are mine alone.

The Department of Artificial Intelligence provided equipment and facilities for my postgraduate studies, even for a fourth year. Jean Bunten earns a thank you for enlivening the department social life, and Olga for keeping such a neat and attractive library. Thanks to all involved in the transfer of the thesis from Edinburgh to Finland - it got there in the end - and also to Judy Delin for handling the thesis on its return. The Department of Linguistics have been marvellous for providing cross-disciplinary and department support. They provided equipment and technicians for two solid days of experimental data collection, and the facilities for the intensive audio-visual transcription. Stuart and Cedrick's frantic assistance as technicians for the study is gratefully acknowledged. Thanks to all the subjects (about 50 people) for taking part in the pilot and the main experiments, and to the Department of AI who provided the money for paying them. Latterly, the English Department of Oulu University has supplied equipment guaranteeing a stable base from which to complete the thesis. The first three years of this thesis were supported financially by the SERC of Great Britain.

Some friends deserve credit for their tremendous support during the ups and downs of thesis life. Flatmates Mark Wiegand and Steve Torchinsky put up with a lot as we all battled on with our PhDs. In the early days of uncompromising lust for the truth, Mike Malloch and Helen Fraser provided me with much intellectual heat as we each wrangled with alternative perspectives to the orthodoxy. Bruce Connell and Anna Gavarro were fine buddies, and together we formed an almost daily mutual support group. Till we meet again in 1999! Thanks to Colin Williams for putting up such a barrage of criticism while defending my ethnomethodological attack on AI, and to fellow PhD peers for griping sessions. The Improvised Music Group proved a useful diversion from thesis work and the motley crowd provided interesting dialogues on the practice of improvisation.

I am most indebted to Pirkko Raudaskoski who has given love, care and sympathy for almost all the downs and ups, as well as providing valuable criticism of my ideas.

Anyway, it's finished now... maybe now I can do that ethnomethodological thesis on how people manage to do a thesis.

I dedicate this thesis to my mother

Jane Tucker

and to my grandmother

Ruth Mitchell

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

> *Not only does the movie and television screen reinforce, by the very velocity of its image and sound presentation, our preconception of past, present, and future in a single line, but only the most sophisticated are aware of the coercion of the technology which prepares the record.... It is extraordinarily difficult to be constantly alert to the extent of control exerted by the focus and the selection of the camera-man and his recording team. Closeups feel right to the experienced viewers, the shift of camera from speaker to auditor, or from speaker to speaker seems natural, too. They influence all of us trained by Western and, particularly, American dramaturgical conventions which see communication, the interpersonal situation, and interaction itself, as action-reaction sequences.* (Birdwhistle 1971).

The term 'dialogue' has a long history and a varied range of senses. Consider the following examples from a dictionary definition:

1. a conversation between two or more people,
2. an exchange of opinions on a particular subject,
3. the lines spoken by characters in drama or fiction,
4. a political discussion between representatives of two groups.

There are Plato's philosophical dialogues, the dialogues of Shakespeare, the lack of dialogue between the United States and the Soviet Union in the cold war, the dialogue debates of political opponents in public, and the dialogues of friends at home over a coffee. The practice of dialogue itself of course has a longer history, indeed as long as humanity. But recently, since the Second World War, the term has appeared in connection with machines and computers. The application can vary from the vision of a conversational, talking computer in dialogue 'with' a human being, to a description of that part of the computer interface that mediates the user's actions through the machine. In addition, what would count as an adequate explanation or description of human dialogue itself has taken a turn both in its conception in a theory and in the techniques adequate for demonstrating a point or argument.

The topic of this thesis is the notion of dialogue and how it has been adopted, changed or coerced from a mechanistic point of view, ie. the technology, metaphors and models associated with machines and their behaviour have led to a largely uncritical adoption of a range of ideas about human conduct that need to be examined to see if they must be rejected, adapted or enhanced in some way. In particular, it is argued here that the interchange model that regards dialogue as the recurrent exchange of discrete messages between individual cognitive processors must be abandoned. Instead, the focus is on the relevance of an interactional perspective that has largely been taken for granted or dismissed from computer models, and therefore from the design and conception of computer systems supporting the dialogue metaphor. Dialogue will be understood to be the collection of practices in which participants are mutually engaged in coordinating communicative actions and achieving shared understanding out of the materials at hand.

Besides the influence of the machine on models of human dialogue activity, recent work in artificial intelligence has suggested conceptions of what role the machine could play in terms of supporting or engaging in human practice itself. It is important to evaluate how successful this research will be, what is required, and what consequences there are, because answers to these questions reflect on what we understand about human conduct, how it works, why it is the way it is, and what forms it could take. Three aspects of 'dialogue and the machine' will be considered at different times throughout the thesis: first, dialogue 'through' machines, which is the use of the computer to support or mediate communicative activities between people; second, dialogue 'with' machines, which is a common metaphor in artificial intelligence proposing that computer systems engage as partners with people in activities that could be called dialogic, and thus the machine's behaviour is to be understood and interpreted in terms of what people do; third, a logical extension is to dialogue 'between' machines, which posits a number of machines engaging in forms of interactive behaviour that might be compared to human communicative activity.

The thesis begins with a consideration of models of dialogue that use computer techniques to explore assumptions about dialogue through the internal representations and programs that can be interpreted to generate simulacrums of the behaviour of people. Current research on discourse and dialogue for complex natural language systems in artificial intelligence has largely drawn upon philosophical and conceptual approaches to language as action or extended linguistic models. However the influence of the human sciences, especially social theory and research, has not been felt. Benefits can be gained by drawing upon empirical studies of language and social action that balance the cognitive emphasis and thus focus on dialogue as an interactive, joint process with unavoidable but crucial local emergent social properties or dynamics. It is argued that abstract computer models of dialogue are too

theoretical and ignore the interactive circumstances of situated dialogue. One particular representative approach, that of Grosz & Sidner (1986), will be examined in detail. It is argued that it has a weak theoretical background and a poor empirical methodology. Also, the interchange model implicit in such computer models may bias the design of natural language dialogue systems and lead theorists to over-rationalise interactional phenomena as the product of an intentional cognitive machinery or ignore them completely. Dialogue systems will then be denied the rich dynamics of interactive dialogue and the development of efficacious systems will be severely retarded.

In order to recover a position that provides a basis for reconstructing more suitable models of dialogue, we need to examine the assumptions about how modelling should be done and justified, what human behaviour is like, and what it is that machines can do. Some terms will be explained briefly at this point to illustrate the nature of the reconstruction. This will also clarify some of the aims and fundamental assumptions of the thesis.

The social objects of interest are the machine and computer, which are often portrayed as mindless, emotionless, hard reasoning and intensely logical, and capable, in principle, even of intelligence and language use. Machines have also influenced how we work and think, and, as Woolgar (1987, p. 325) states, "discussions about technology embody fundamental preconceptions about the character of mankind." Two uses of the machine are of special interest here: first, the **dialogue artifact**, which is a computer whose use is premised on the metaphor of human dialogic communication and is embedded in human practice and intended for human consumption; second, the **virtual dialogue space**, which is a computer-simulated graphic space in which people can participate as if in a dialogue.

The starting point of the thesis comes from **ethnomethodology**, **conversation analysis** and **hermeneutics**, all of which study interpretation and the social construction of intelligible action in human conduct. Unfortunately they are difficult fields to understand and for a number of reasons: they are from a completely different paradigm of explanation, they are often published in a foreign language and translations are late, the style of writing is found difficult, and the theories are radically different from the rationalist tradition of theorising about human conduct exemplified by Chomsky, Descartes and Plato. But once understood, these fields provide the basis for a reconception of the nature of the machine in human practice, for an empirical methodology for the analysis of fleeting conduct, and for a radically different idea about action and communication. To expand on the latter idea, **situated action** is a term used to suggest that action is always situated, always comes out of some set of physical, social and temporal circumstances and is contingently achieved by those engaged out of the materials that make up that activity. This is in marked contrast to the view of action and dialogue supported by computer models that favour the exchange and recognition of intended

messages between individuals. Instead, shared understanding is a local achievement that is not the result of instantiating social norms or the playing out of cognitive rules. In the following quotation, Schegloff sums up the situated character of communicative action, and argues clearly that a notion of interactivity is essential to the investigation of discourse:

> *Good analysis retains a sense of the actual as an achievement from among possibilities; it retains a lively sense of the contingency of real things. It is worth an alert, therefore, that too easy a notion of 'discourse' can lose us that. If certain stable forms appear to emerge or recur in talk, they should be understood as an orderliness wrested by the participants from interactional contingency, rather than as automatic products of standardized plans. Form, one might say, is also the distillate of action and/in interaction, not only its blueprint. If that is so, then the description of forms of behaviour, of forms of discourse (such as stories included), has to include interaction among their constitutive domains, and not just as the stage on which scripts written in the mind are played out.* (1982, p. 89)

The thrust of the present thesis is to demonstrate that an interactional perspective on dialogue can help us come to an understanding of the problems of computer models of dialogue and a conception of the dialogue artifact. A step towards this is made through an empirical study of the circumstantial and interactive achievement of dialogue 'through' machines. It is the aim of the thesis to demonstrate that not only the description but also the modelling, simulation, and practical generation of human-like dialogue conduct must take into account interaction as a constitutive domain of dialogue and cognition. In order to do this, hermeneutics and the empirical methods and insights of ethnomethodology and conversation analysis are used to investigate problems with computational theories of dialogue and action as well as to understand the nature of interactive artifacts and their future design and use.

Human dialogue has been studied by many disciplines, but one of the most successful accounts of everyday dialogue - conversation analysis - has explained a number of sequential and interactional details of conversation and institutional talk. In contrast, the thesis considers a particular example of machine mediations of human dialogue in which dialogue 'through' the machine is achieved. An empirical audio-visual study of computer-mediated dialogue forms the basis for revealing features of the taken for granted achievement of dialogue in restricted contexts and for testing the applicability of a methodology derived from conversation analysis. The primary focus is on the methods and resources, usually taken for granted, that people use for achieving and maintaining the 'sense' of their talk and actions and for displaying that 'sense' to each other. Techniques are used for engendering accounts of the engagement in dialogue and a complex transcription system is developed to render the detailed work of constructing contributions in the dialogue. The analyses focus on the achievement of presence and participation. Action is constructed in and through technology at hand, and its use is shaped by the demands of the activity. Thus inscriptions in the virtual

dialogue space are only sensical in the context of the practices that gave rise to them. Also, participants must monitor their own and each other's actions and construct their participation from the materials provided for in the virtual dialogue space. Because of this, the model of conversational turn-taking is undermined and new orders of participation emerge. A second, brief study looks at how the collaborative work of two people is finely coordinated through the embodied performance of action and the complimentarity of actions by the same or more than one person.

From the analyses it is concluded that a wealth of resources and constraints are contained in the taken for granted work of dialogue activity. It is also concluded that dialogue activity is shaped by circumstances such that the order of participation is transformed unintuitively. Thus the interchange model which takes no account of the circumstantial, situated and interactional constitution of dialogue is inadequate and must be abandoned. Dialogue is more than the abstract exchange of discrete messages between entities. This means that the following three consequences of the model must be acknowledged and dealt with: the lack of self-coordination and repair, the lack of recipient design in concurrency, and the strict interchange of participation. It is recommended that the findings of interaction analysis be incorporated into the design of artifacts supporting the metaphor of dialogue 'with' machines. In particular, dialogic presence and local participation are features of mutual engagement that can be designed for with careful evaluation. It may be possible to achieve a satisfactory user and machine engagement from discrete materials, and avoid the time-consuming, computationally inefficient, and over-representational procedures characteristic of computer artifacts in artificial intelligence. The success of the interaction analysis methodology for exposing the situated work of dialogue in reduced contexts of presence means that the methods and findings may be useful in general for studying and evaluating dialogue 'through' and 'with' machines. A major consequence of the arguments in the thesis is that traditional computer modelling techniques may not reveal an understanding of how people engage in activities such as dialogue and thus a reconstruction of the nature of modelling is proposed briefly for future work.

The linear structure of the thesis will be briefly summarised. In Chapter 2 the development of computer models of human dialogue is traced from its origins in traditional models of dialogue, computer modelling and natural language processing. The chapter demonstrates the problems and limitations of such computer models and motivates an alternative interactional perspective that is expanded in Chapter 3. One particular approach, that of Grosz & Sidner (1986), is examined closely in terms of its theory, empirical methodology, explanatory power and implementation criteria. Three major points made in this chapter are: that computer models of dialogue such as Grosz & Sidner's have a weak theory of action

and activity, that the empirical methodology leads to a sanitised view of dialogue, and that such computer models provide a poor foundation for the conception and design of computer dialogue systems supporting the metaphor of dialogue 'with' a user.

Following on from the critique, Chapter 3 develops an interactional perspective on the nature of dialogue and an understanding of the machine in use. First, it is important that the conception of the machine and its behaviour be examined in order to understand in what ways the machine can be considered to take part in practical dialogue. Drawing from hermeneutics, the notion of the machine as an intelligible artifact is discussed in contrast to the notion of the 'intelligent' machine in artificial intelligence. Next, the field of ethnomethodology is described because it provides an alternative conception of the practice of activities such as dialogue. Rather than reconstruct dialogic interaction as the recognition and coordination of plans, the specific mutual achievement of dialogue in situ - from which rational accounts emerge - is studied. From the field of conversation analysis, dialogue is understood to be a collection of practices in which participants are mutually engaged in coordinating communicative action and achieving shared understanding in situ. The major principles and naturalistic methods of these two fields are explained and illustrated. From the review and arguments above, an interactional perspective is derived that remedies some of the major problems brought out in Chapter 2.

An investigative audio-visual study of dialogue 'through' the machine or computer-mediated dialogue is described in Chapter 4. The study examines the achievement of dialogue between people when it is mediated through the use of the computer. It was carried out to reveal some of the situated and circumstantial details of dialogue that are taken for granted or ignored in computer models of dialogue; to demonstrate the applicability of an approach derived from conversation analysis to computer-mediated communication; and to recommend resources and environments for computer dialogue artifacts. A brief overview of the study is given, and a survey of related research brings out contrasting principles and techniques to those used here. The principles and procedures of the study are described in depth. Techniques of recording and transcription were developed specifically for this type of dialogue setting. Finally, the complex records that form the corpus for the analyses in Chapters 5 and 6 are explained and illustrated. The complete set of notation conventions are in Appendix A, and some of the longer examples are to be found in Appendix B.

Chapter 5 presents a detailed analysis of the situated, interactional achievement of dialogue 'through' the machine in a virtual dialogue space. Communication is a skilled accomplishment wrested by the parties from the circumstances and emergent dynamics of their mutual engagement. First the chapter argues why the screen display and a turn-based rendering of the data are inadequate for analytic purposes, and then a variety of interactional

concepts relevant for the analyses are presented. The analyses document with examples: the situated work of constructing action in and through the technology, the visual and temporal monitoring of conduct, the local management of participation, the relevance of contributions to each other in dialogue, and the emergence of interesting new orders of participation constructed out of the materials at hand. In conclusion the relevance of the findings to computer models of dialogue is discussed.

The study reported in Chapters 4 and 5 demonstrates that an interactional analysis can reveal the situated, interactional work of constituting dialogue 'through' machines, and Chapter 6 considers the consequences of the methodology and findings for dialogue 'with' machines. It is argued that features of dialogic presence and interactivity are essential for the design of more appropriate computer dialogue artifacts, some of which have been restricted by the interchange model. Also, a brief study of copresent collaborative work in a workspace is presented that documents two other features of interactivity: the embodiment and complementarity of actions. Thus the thesis documents a range of features, if designing for interactivity, that are both valuable constraints on interpretation and resources for constructing intelligible action. However, much more careful design and evaluation is needed if an appearance of 'interactive flow' is to be achieved in which participation in dialogue is itself locally constructed and artifacts are maximally sensitive to the user's participatory needs.

The concluding chapter discusses a number of issues arising from the preceding chapters. The criticisms of computer models of dialogue are considered in the light of the findings of the thesis, and prospects for incorporating the insights into modelling are discussed. However, the interactional perspective suggests a more radical reconstruction of the way modelling could be carried out, so the use of computers to model dialogue activities is returned to at a deeper level in order to reconstruct an alternative modelling paradigm for investigating the emergence of embodied and dynamic activity. A brief summary of 'embodied modelling' is given that emphasizes the emergence of interactional dynamics and aims to realise interesting communicative activities 'between' machines. The empirical methodology is looked at reflectively to see what improvements could be made in hindsight, and what further work could be undertaken. Much further research is needed to investigate how people coordinate their mutual activities with body, speech and artifacts. Further applications of the findings and interactional perspective are discussed in relation to dialogue 'with' and 'through' machines. Possible extensions to the interaction analytic methods are speculated on before general considerations about the nature of dialogue and the machine are briefly discussed.

Chapter 2

COMPUTER MODELS OF DIALOGUE: THEORY, DATA AND IMPLEMENTATION

2.1 Introduction

It is appropriate, first of all, to consider how conceptions of dialogue in research have been influenced by the machine and mechanistic notions of behaviour in order to consider in turn how an understanding of dialogue has influenced the design of machines intended to engage in dialogue 'with' people. Of course, this is not a natural ordering - ideas about human behaviour and the development of technology have gone hand in hand. However, currently it is claimed that computer models of dialogue can help designers of 'intelligent' interactive computer systems, so it is important to know just what the assumptions and limitations of these models are. This chapter sets out to demonstrate the problems and limitations of computer models of dialogue similar to Grosz & Sidner (1986) and to suggest an interactional perspective to supplement or supplant the weak notion of dialogue in these models. In particular, it is argued here that the traditional understanding of dialogue in artificial intelligence and cognitive science is derived from an individualistic, procedural account of action that is over-rationalised and misconceived. The assumptions that coherence and order is based upon intent recognition, that there are conventions for the expression of intent, and that there is an enumerable body of shared background knowledge are challenged. Because of the weak empirical methodology, theories and models are allowed to develop without due care for the interactional, situated and circumstantial details that are endogenous to the achievement of action and dialogue activity. The fictional treatment of dialogue examples in the documentation and analysis of intent ascription shows a disregard for the circumstantial production of dialogue, and thus the interchange model of dialogue is reproduced. Consequently, if models are applied to the conception and design of dialogue artifacts then impoverished performance will result, for example, inflexibility in situated

contexts or poor opportunities for the user to 'interact' with the system at appropriate moments. Solutions to the problems that arise in performance, or that are hypothetically raised by the analyst intuitively considering the possibilities, are then suggested by those same theories resulting in a closed consideration of internal problems that do not address the fundamental issues. An attempt to rethink the conception of dialogue and the machine is undertaken in this thesis.

Dialogue will be understood as mutual activity involving multiple participants engaged in practical coordination and communication using natural language, eg. correspondence, conversation, debates, etc.. The focus of the chapter will be on research that studies how dialogue can be interpreted or generated in practice. A brief survey of work on traditional models of dialogue, computer modelling and natural language processing is described in order to trace the developments leading towards the investigation, using computational tools and a mentalist vocabulary, of communicative action in dialogue. The focus is on those models that turn away from the problem of describing a phenomenon as a structured product to the issue of how individuals produce or interpret the phenomenon. Another concern is with models that would be useful for conceptualising or building a computer dialogue system. Unfortunately most work is primarily concerned with written language use and spoken language use to a lesser extent, with little research on the non-verbal and situational characteristics of communicative action.

The main classes of computer model are described and then one particular computer model of dialogue is examined in detail in terms of its theory, empirical methodology and implementation. In Grosz & Sidner (1986) a theory is described that tries to account for purposeful discourse activity, discourse structure and the processing of discourse contributions. This model has advantages over other models, but there are limitations when dialogue is considered with respect to a theory of situated interactivity and an empirical perspective that would provide much evidence for theory development. For example, the evidence that the treatment a contribution gets in dialogue is not static, but can be reclassified or misplaced, must point to a reconsideration of the integrity of the segment and the segmentation procedure. The intention here is not to develop or replace their model with another, but to indicate the problems and resources available for alternative directions that may supplement or supplant the class of similar models. The remainder of the thesis will then examine empirically the constitutive domain of interaction in the context of a particular study of computer-mediated dialogue, and consider the development of appropriate methodologies for designing and evaluating useful computer dialogue systems - dialogue 'with' machines - and for exploring the nature and emergence of situated dialogue activity.

2.2 A Short History of Traditional Models of Dialogue

To trace the development of traditional theories of dialogue within the extraordinarily diverse and rich field of discourse analysis would be a difficult task[1]. It is not undertaken in this thesis. But, a few key developments and issues must be mentioned to understand the recent interest in computer models. Useful guides can be found in Coulthard (1985/77), Stubbs (1983), Brown & Yule (1983), Levinson (1983), and Dijk (1985) that consider dialogue as well as other forms of discourse, eg. monologues. Dascal (1985) is a multidisciplinary collection based on the theme of human dialogue that acknowledges that the phenomenon of dialogue, a natural human activity, is a major research topic.

Research on dialogue was fragmented up until the second world war and confined to that which was considered to be historically, artistically or rhetorically momentous or important. Indeed, everyday language use was largely ignored and dismissed. Since then work has proliferated and progressed on two levels - the utterance and discourse level - within the fields of psychology, linguistics, philosophy of language, and sociology among others. The methodology adopted has varied from the deeply empirical and descriptive to the intuitive and constructed. One of the first formal linguistic approaches by Harris (1952) proposed a linguistic model "for the analysis of connected speech and writing" that attempted to apply distributional techniques to discourse. Since then many models have been proposed, for example, the exchange structure model of the Birmingham discourse school described in Coulthard (1985/77) which analysed institutional dialogues in order to develop a structural analysis in terms of the realisation of units of discourse in the manner of a grammatical theory. Philosophy has also had a great influence on theories of dialogue that attempt to ground language in action and cooperation. For example, speech act theory was developed by Austin (1962) and Searle (1969), but is not really concerned with dialogue, only the conditions required for the successful achievement of a speech act by a speaker. This approach has been heavily criticised recently: for example, Levinson (1981a and b) argues that speech act theory is problematic as a model of dialogue. A truly dialogic analysis came from Grice (1975) who examined the cooperative basis for conversation in terms of conventional conversational maxims and the logic of the inferences or implicatures that can be made by speakers and hearers in conversation. These studies come under the banner of pragmatics. A radically different approach, conversation analysis, also emerged at this time to study empirically the

1 The terminology is a great problem as well. For instance, dialogue, discourse, and text are not cognate concepts.

interactive achievement of natural everyday conversations. This development, which informs the remainder of the thesis, will be discussed later in Chapter 3.

Debates have been furious across the fields that uneasily congregate under the flag of discourse analysis. Multidisciplinary perspectives unfortunately have led to paradigm clashes and misunderstandings. In particular, a key concern has been with the nature of the order to be found in dialogue. Structural notions - eg. move, turn, interchange, section, lesson, exchange, conversation, story - have been introduced in order to explain an order of dialogue amenable to analysis or categorisation. In many approaches these structures are posited as knowledge. However, the status of that knowledge in the dialogue itself and how that order is produced is unclear. For example, are there rules that govern dialogue conduct? These issues lead into the most relevant development for considering computer models and that is the emergence of a process or procedural perspective.

Brown & Yule (1983, p. 23-4) note three distinct stages in the field of discourse analysis. First, the sentence-as-object view, which is mainly descriptive and includes most of the field of linguistics. Second, the text-as-product view, which looks at the discourse textual record as produced by a speaker in a context, but without regard to how the product is produced or interpreted. And lastly, the discourse-as-process view which is summarised as follows.

> *We shall consider words, phrases and sentences which appear in the textual record of a discourse to be evidence of an attempt by a producer (speaker/writer) to communicate his message to a recipient (hearer/reader). We shall be particularly interested in discussing how a recipient might come to comprehend the producer's message on a particular occasion, and how the requirements of the particular recipient(s), in definable circumstances, influence the organisation of the producer's discourse.* (ibid, p. 24)

Discourse is:

> *a dynamic process in which language was used as an instrument of communication in a context by a speaker/writer to express meanings and achieve intentions (discourse). Working from this data, the analyst seeks to describe regularities in the linguistic realisations used by people to communicate those meanings and intentions.* (ibid, p. 26)

The essence of this third development is that the production or comprehension of discourse is itself a constitutive domain of discourse organisation that must be studied. For example, a schema-based approach that emphasizes data structure schemata in the process of comprehension of stories is described in Rumelhart (1975). He claims that the theory embodies not only a general theory of meaning but a procedural one as well. In other words, the procedures implemented in the model parallel the procedures by which people comprehend discourse. This approach was part of the developing cognitive science perspective to language and cognitive phenomena that uses the computer metaphor in its

study of mentalist processes like dialogue reasoning. (See Cicourel (1980) for a critique of this type of model as well as non-procedural models.)

To summarise, there are many approaches within discourse analysis and there is no attempt here to bring them into a unified perspective. What is important is the contrast between the descriptive tradition and the newer interest in explanation or description through simulation - or as Levinson (1981b, p. 487) puts it, "analysis-by-synthesis". Moreover, the process view traced and outlined above has come to mean more than describing the regularities in discourse as the product of a linguistic process. For instance, Cohen (1984) discusses the work of Labov & Fanshel (1977), a standard approach in discourse analysis, and criticises the extent of their claims:

> *Their analyses presented rules for interpreting the intentions behind utterances of various syntactic forms - eg. rules for when a hearer will interpret utterances as indirect requests for physical action or verbal confirmation. However, these rules were stipulated as regularities of discourse rather than as derived from underlying processes. Their findings should serve as data to be explained, rather than as a satisfying account of discourse.* (p. 100)

2.3 A Background to Natural language Processing and Computer Modelling

Before looking at recent work in computer models of dialogue, a summary of the principles of computer models of language phenomena is needed. The use of computers in the study of language has given rise to the field of natural language processing in which most work is on grammar in the domains of morphology, syntax, lexis, etc.. Winograd (1983), Grosz et al (1986), and Spark-Jones & Wilks (1983) contain useful summaries and readings of the developments in the field. For various reasons, the development of complex computer models of interactive discourse has been slow compared to the intensive work on intra-sentential language structure, semantics and text generation. It seems probable that this was in part a result of the work on computer languages in the 1950s, combined with the rise of Chomskyan linguistics, which led to parsers that can procedurally categorise strings into constituent structures. However, the important point is that the confluence of ideas from computer science and linguistics led to a quite different conception of language and how it should be studied.

The computational metaphor has been adopted in different forms by many disciplines. Winograd & Flores (1986) summarise the main assumptions of the 'computational metaphor' in terms of cognitive systems:

> *1. All cognitive systems are symbol systems. They achieve their intelligence by symbolising external and internal situations and events, and by manipulating those symbols. 2. All cognitive systems share a basic underlying set of symbol*

> *manipulating processes. 3. A theory of cognition can be couched as a program in an appropriate symbolic fashion such that the program when run in the appropriate environment will produce the observed behaviour.* (p. 25)

The metaphor is also deeply embedded in how human communication is theoretically understood, and how artifacts should be conceived and designed. The standard model is summarised in Winograd (1983, p. 12-3):

> *Language is a process of communication between intelligent processors, in which the producer and the comprehender perform complex cognitive operations. The producer begins with communicative goals, including effects to be achieved, information to be conveyed, and attitudes to be expressed.... In order to communicate the producer must map this multidimensional collection of goals onto a sequence of sounds that can be uttered or marks that can be drawn on a page.... The choice of words, the structure of phrases, and the patterns of emphasis and intonation all play a part in providing the necessary cues for the comprehender to infer the producer's goals.*

What the metaphor has given researchers is rigour in the specification of the theory in the form of a computer model. Any procedure must be computationally decidable, ie. it must be possible to compute the procedure in finite time, and also tractable, ie. computable in reasonable time and resources. But the exact status of the computer program - its behaviour, procedures and symbolic representations - to a theory is unclear, thus computational rigour is not necessarily useful as proof of the 'correctness' of a theory. Anyway, provided explicit accounts of the procedures are given, the rigour of implementation also supports a paradigm principle similar to that found in the physical sciences, namely reproducibility. Another person should be able to reproduce the machine behaviour or results in another place and on a different machine from the formal specification of the computational procedure.

With variants of this metaphor, computer models have formed into three general types. The success of any account is to be judged in terms of the following kinds of adequacy:

> i) Descriptive: allows extrapolation of assumptions by the computer generation of descriptive linguistic regularities. This is common in computational linguistics.
>
> ii) Generative: simulation of human-like language behaviour in artificial intelligence.
>
> iii) Cognitive reality: a postulated computational mechanism faithful to human mental language processes without having to specify neurological or biological details. This is the basis of most work on language in cognitive science.

In account i) the aim might be to produce a macro model that describes or categorises the global properties of language or discourse, or a discourse process description that predicts the referent of an anaphoric pronoun in a discourse script. A primary aim is to propose

procedures for mapping one representation onto another, but there is no claim for adequate simulation of behaviour or cognitive adequacy. The next two measures of adequacy concern relations between representations on the one hand, and the world and action on the other. Account ii) requires the engineering of computational processes in order to build 'intelligent' systems with some exploitation of an understanding of human language conduct, but that also incorporates a "do it anyhow" method, eg. plan-based approaches to intelligent interface design, discourse models, user models, history models, quantitative methods, etc.. This may be achieved with no resemblance to postulated human subpersonal mechanisms. Also, it may result in real artifacts that are designed to engage with people in similar activities to those modelled. The operational Turing test is a somewhat old characterisation of machine intelligence appropriate to the aims of much of this work (Turing 1950). It is the enterprise of simulating or realising appropriate behaviours that is most widely considered here. In account iii) psychological validity places an additional set of restrictions because production and generation mechanisms and representations must all be justified in mentalist terms, with the hope that neurological and biological theories will eventually support them.

What about the origin of examples of the phenomena to be explained and evidence for the theory, and the status or use of computational procedural models? Examples of the phenomena are usually derived from intuition or verbal records of restricted experimental interactions. An additional source of data is from the implementation, which is a testing ground for behavioural aspects of the theory. Assumptions can be tested within the model's limits by running long computations that cannot be done by hand. But there are problems with many implementations, especially when one considers the practical generation of human-like language use. Implementations tend to have the following characteristics: pre-constituted inputs and outputs; no real practical consequence or situation; human interpreted and assisted process; hypothetical scenarios and results; and unlimited computation time and space. For example, an utterance in a hypothetical discourse may be preconstituted as a grammatically correct string of words to be input in written form to a computer program that will output a similar form to be judged by a fellow analyst as the next utterance in the hypothetical discourse. Also, plan-based systems, for example systems such as described in Appelt (1985), based on intentional models of language generation and action can take several minutes or even hours to compute a next linguistic action.

The computational metaphor has been subject to many criticisms, eg. the inadequacy of representation as a means of producing intelligible conduct through computational processes. Even Winograd, an early proponent of the metaphor, has himself reflected and altered his position in Winograd & Flores (1986). They place the work of AI within the rationalist tradition derived from the physical sciences and pervasive in philosophy since the

times of the Greek philosopher Plato. Emphasis in this tradition is placed on the formulation of identifiable well-defined objects and formal rules which operate on the objects, such as in symbolic logic or automata theory. The particular problems for achieving an appropriate model of dialogue are to be found in the tendency to align with the rationalist tradition applied to the study of human interpretation and understanding. The discursive and conceptual explanation is presumed to be above the practical and situated, and consequently there is a tendency to reify the explanation in the head of an individual, ie. as the means by which conduct is generated. But, because of the requirement for computer implementation, problems are exposed in the less than perfect performance of the artifact, designed on the basis of a theory, in practical circumstances with human users. This has been shown by Suchman (1987) in the case of an expert help system, based upon plan-based theories of action, for users of a photocopier.

However, besides the complex of difficulties, a progressive part of computational accounts is the move towards an examination of the relations between structural models and the processes of production and interpretation, ie. descriptions and intuitions about language structures vs. the real-time process of producing and interpreting action in the world that gives rise to that structure. Much research is guided by the hope of building and subsequently evaluating machine artifacts that can act intelligently and communicate with humans.

To conclude, given that computational accounts of language and action have moved from a 'product' to a 'process' view, computational theories claiming an abstract account of the process of conduct have been reticent about the actual details of situated practices. Usually, an account is formed from our folk notions about the nature of conduct, and then some process or mechanism computing over representations is hypothesized in order to generate the behaviour or regularities. This computational process is then reified as the cognitive process or as some abstraction from it. Of course, this work is useful and productive, in that interesting computer systems for real use may emerge as the result of these investigations, as well as theoretically stimulating, in the sense that it may suggest through its efforts the impossibility of building an intelligent computer. The weakness is that it is methodologically individualistic and lacks external constraint, which consequently leads to over-representation and reification.

2.4 Computational Accounts of Dialogue: Theories and Models

2.4.1 A Survey

Much current computational research on language is concerned with generation and interpretation, particularly of single speaker/writer sentences or utterances. It is, however, gradually expanding into the study of multi-sentential discourse. This thesis is only concerned with approaches to dialogue that attempt to deal with mutual intelligibility and the collaborative achievement of dialogue by participants. That is, the concern is with accounts in which a response by one party is addressed in the next sequential slot by another party, and so a response must be generated in real, practical circumstances.

First let us consider the early artificial intelligence (AI) systems that were claimed to be demonstrations that machines can 'understand' human language in dialogue. Programs simulating simple dialogue behaviour emerged parallel with the development of interactive computers. Question-answering interfaces to database systems will not be considered because they do not attempt to provide coherent linguistic messages over extended exchanges with the user, but they are still an important link in the development of computer models of language use. Instead, consider the infamous ELIZA program, Weizenbaum (1966), that used simple techniques to achieve a simulation of a Rogerian therapist. This was lucky for the programmer, because the fundamental principle of this form of psychiatric therapy is to remain passive and reflect problems back to the patient for them to discursively mull over with therapeutic consequences. The program is quite successful in achieving a reasonable dialogue; or rather, the users are quite successful at finding the language output intelligible as contributions in a dialogue. In fact, for limited periods people can believe that they are engaged in intelligent dialogue 'with' a machine. However, it does not take long for the program to inadvertently slip up and thus demonstrate also its simplicity. Several similar systems developed, but after the initial excitement it was realised that something more was required than just a clever set of procedures exploiting the propensity of people to find sense in almost anything (as Garfinkel (1967) has demonstrated) and certain stable forms of language that occur in particular activities. McTear (1987, p. 67) summarises the simple techniques that were used in these systems:

1. The domain was highly restricted,
2. The subset of natural language input was restricted,
3. A keyword matching procedure was used,
4. The input was mapped onto an underlying artificial/formal language.

A major step forward came with Winograd's SHRDLU which was supposed to answer restricted questions about the state of a simple microworld consisting of different coloured toy blocks, to perform the user's linguistic commands with the aid of a hypothetical robot arm, and to gain knowledge from statements by the user (Winograd 1972). This was a big advance on the earlier template-matching programs because of the use of syntactic and semantic theories to design the procedures and modules of the program. Even though the robot was passive, Winograd explored some of the key problems of how language is used in multi-utterance dialogue in which utterances have to be interpreted in the context of physical objects and activity.

Since this period, various attempts at simulating human dialogue in restricted task domains have been made, eg. Bobrow et al (1977). But, in common with cognitive theories of discourse processing, there has been a separation into both real system development and focused research on particular modules that may play a part in the generation or interpretation process. This work has led to the publication of several collections of research on computer modelling of dialogue, eg. Joshi et al (1981) and Brady & Berwick (1983). In Joshi et al (1981, p. 1) the following four features are stated as necessary parts of a model accounting for discourse:

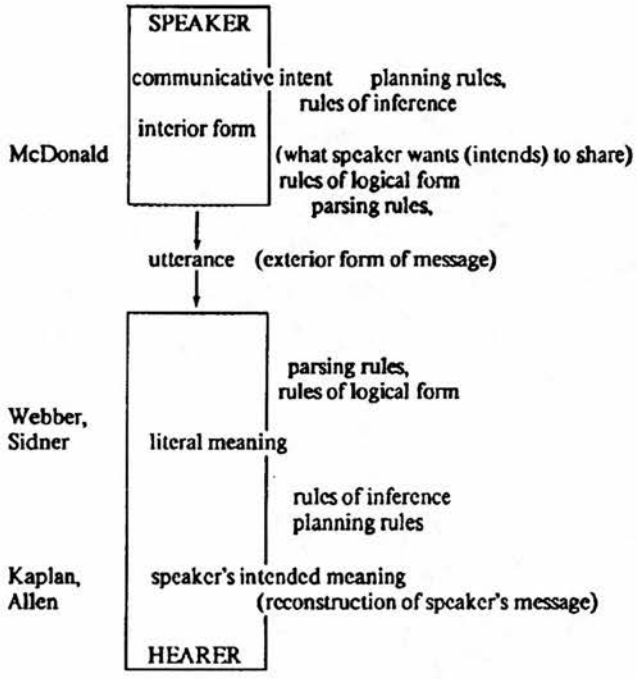
a. Utterance meaning,
b. The discourse participant's evolving model of what underlies the discourse,
c. Their similarly evolving models of each other, user modelling,
d. Situational characteristics.

The over-riding concern within the artificial intelligence tradition since then has been to specify what sorts of 'knowledge' are required in order to interpret or generate a discourse contribution using the model of language use shown in Figure 2-1.

The specification of knowledge in terms of user or discourse models is considered in Kobsa & Wahlster (1988a and b). In the proliferation of knowledge-rich approaches, two lines can be distinguished: those that use language structure and those that look outside the linguistic context at the mental state or activity that a participant is engaged in. Adherents of the possibility of a linguistic explanation or model of dialogue come mainly from linguistics and adherents of the over-riding importance of 'world knowledge' are from philosophical and psychological domains, among others.

Just before looking in general at linguistic and non-linguistic models, the type of data used in computer models and its status needs clarification. Evidence of language use in dialogue is usually intuitive or constructed for the concerns of the theory. This is very much in the style of speech act theory and traditional linguistic analysis. Examples of plan-based or rational

FIGURE 2-1 - A model of language use in dialogue from Brady & Berwick (1983).



theories of speech action are to be found in Cohen & Levesque (1988), and in most of the work in Joshi et al (1981) and in Kobsa et al (1988a). Some data derived from human dialogue is obtained from experiments, eg. Cohen (1984), or naturally recorded, eg. Reichman (1985), but the data is poorly rendered for analysis with little attention to the local circumstances of interaction. The examples are then used as the input to an 'understanding' model to categorise, segment or make valid inferences from on the basis that they are representations of coherent utterances in a dialogue. One could say that the results of a 'generation' implementation are data, too. But, in many cases implementation is not fully achievable, and thus idealisations of data sets must be used in order to gloss over the parts of a complete model that are not implemented. For example, the difficult problem of speech perception is avoided by working with decontextualised utterance objects made up of strings of characters forming words. Also, the evaluation of the performance of an implemented model as a 'dialogue partner' is limited, usually being at a conceptual level of agreement by analysts over whether or not a particular structure or 'interpretation' could be assigned or an appropriate response given to a speaker's utterance in a constructed or cleaned-up real dialogue, eg. Walker(1989). We shall see examples of this later in the critical consideration of Grosz & Sidner's data and their implementation criteria.

2.4.2 Linguistic Approaches

At one end of the spectrum, Polyani (1983) has favoured a strict extension of the linguistic approach to the study of discourse structure. He posits different levels - local structures, discourse units, speech events - with a grammar of each level analogous to linguistic patterning at the sentence level. Other approaches are not so entrenched within a strictly syntactic framework but they do posit some ordering at the linguistic level. The majority of computational models have drawn upon rhetorical units that are deeply embedded in the social action vocabulary of analysts and actors. Some sort of discourse unit is defined and then a recursive embedding structure relating units to each other is hypothesized to explain the structure of discourse. Reichman's approach to 'everyday conversation' is explicitly linguistic (Reichman 1985). She is interested in functional relations between rhetorical conversational moves that are indicated by clue words in dialogue. The ultimate aim is to derive a model grammar that governs our conversation and that can be implemented as a computational procedure reflecting the process of human dialogue. In Reichman-Adar (1984) she acknowledges the abstract nature of the model she presents but argues that to build natural language interfaces one has to be sensitive to the rule-governed nature of extended human communication that her model outlines. Bateman (1986) takes a distinct linguistic stance to the organisation of dialogue derived from systemic linguistics, in the style of the Birmingham discourse school, and incorporates some of the elements of conversation analysis. He explicitly avoids the use of an intentional vocabulary by investigating the discourse functions of linguistic items, and the relations between the deployment of linguistic resources and contextual development.

2.4.3 Non-Linguistic Approaches

The principle adopted in the non-linguistic modelling of dialogue is that the coherence of dialogue cannot be explained at the linguistic level itself but must be the result of an underlying order or structure that is perceived and used by participants in the dialogue. The underlying structure is derived from facts about the world, the activities the dialogue participants are engaged in, or their mental states. For example, the work of Levin & Moore (1977) uses a dialogue game approach that typifies certain distinctive features of the possible activities that dialogue participants could be engaged in - eg. helping, information probing, that are similar to discourse-level speech acts. Dialogue consists of the application of a dialogue type to constrain interpretation and production on the basis of recognition of the speaker's goals and knowledge states identified from an utterance. But, the dialogue types are too restrictive and clumsy, and the model does not have a well developed theory of action.

Conceptual approaches that hypothesize mental states, like intentions and beliefs, of individual actors and that use a theory of language as action that relates language form to performed action have become popular. This is because of the realisation that dialogue is conducted for a reason and participants usually have a purpose in acting the way they do. This work in computational fields has borrowed and developed philosophical accounts of language. For example the early development of this form of computer model tried to combine the speech act model with a plan-based theory of action, eg. Cohen & Perrault (1979). Appelt (1985, p. 1) argues that "human language behaviour is part of a coherent plan of action toward satisfying a speaker's goals" and that speech contributions encode speech acts designed to effect certain actions in the plans of the speaker. Also, Cohen (1984), talking about the plan-based approach to communication, states that:

> *the coherence of dialogue is to be found in the interaction of the conversant's plans. That is, a speaker is regarded as planning his utterances to achieve his goals, which may involve influencing a hearer. On receiving an utterance, the hearer attempts to infer the speaker's goal(s), and to understand how the utterance furthers them. The hearer then adopts new goals, and plans his own utterances to achieve those. A conversation ensues.* (p. 111)

Oviatt (1988, p. 47-51), on cooperative conversation, says:

> *Computer systems that function as cooperative conversants... are based on an approach to communication that treats utterances as actions that a speaker plans and reasons about in attempting to alter a listener's mental state. Such actions have been called speech acts by philosophers of language.... a theory of rational interaction provides the foundation for understanding communication.... In the analysis of cooperative dialogues, it is postulated that listeners are motivated to understand why speakers say what they do, and to infer their communicative goals and plans. It is important to specify that cooperative dialogue ideally involves mutual recognition of the intentions behind one another's utterances.... That is, listeners attempt to infer what the speaker intends to accomplish through communication, rather than adopting a superficial interpretation based exclusively on literal statements and observable events.*

Recently it has been recognised that a more general theory of rational communication is needed in order to understand "autonomous action by rational agents." (Kiss 1988/86, p. 4). Cohen & Levesque (1987) argue that analysis in terms of speech acts is unnecessary because properties of speech acts can be derived from the principles of a general theory of rational interaction. Crucially, they claim that "what speaker's and hearer's need do is to recognise the speaker's intentions" (p. 3) rather than recognise the illocutionary act performed.

This type of model will be examined in greater detail in the next sections. Basically, the conception of interaction and language use is rather narrow, viz. a view of dialogue as the rational exchange of extended utterances by more than one participant in which complete

actions with an underlying intent are recognised by the recipient at the utterance-level and over extended exchanges. One contemporary model that has tried to span both the linguistic and non-linguistic approaches discussed in the last two sections is that of Grosz & Sidner (1986). The remainder of the chapter will take a detailed look at this theory and its computer model of dialogue.

2.5 Grosz and Sidner's Theory of Discourse

So far, a number of computer theories and models have been described that are relevant for an understanding of the processing of dialogue, ie. for the computational representations and processes required for the production and interpretation of adequate dialogue contributions as a participant in that dialogue. Note that the thesis is not immediately concerned with the identification, categorisation and representation of structures to be found in dialogue activity using computer models. Many models only aim to provide such an account and they are beyond the scope of the thesis because the structures may play no role, tacitly or discursively, in the practice of dialogue itself. For instance, participants do not draw upon statistical measures of utterance-type frequency across gender categories in order to construct their intelligible dialogue, though there may be a useful structural pattern to be found by computer analysis. Of course representations of interesting structures may be useful for AI dialogue systems that can use any computationally viable means to generate or interpret an action, though a parsimony with respect to the importation of internal representations and mechanisms is recommended in Chapter 6.

One particular contemporary approach, Grosz & Sidner (1986), has been reserved for careful consideration for a number of reasons. Grosz & Sidner's theory of discourse is a comprehensive attempt to provide a model that relates linguistic structures of dialogue to the non-linguistic processes of production and interpretation. First, the model is concerned with purpose: participants in a dialogue engage in it for a communicative purpose, and this fact informs the construction and coherence of the dialogue. Second, it deals with linguistic structure: that language structure does exist above the sentence, and language form is crucial in the process of dialogue and should be integrated with a non-linguistic model. The third concern is with process: that a locus of focused processing of dialogue contributions occurs over time. Fourth, the model includes activity: the general activities and tasks engaged in by the participants are important. It has a number of improvements over previous theories, including the implicit recognition of the shared nature of discourse structure - "Discourses are fundamentally examples of collaborative behaviour." (Grosz & Sidner 1987, p. 4) - and improvements on rhetorical approaches to discourse structure. As a consequence of this, their theory is taken to be representative of many of the other approaches described earlier.

In addition, their theory concerns communicative action, not specifically verbal dialogue, and so it should apply to a wide range of dialogue contexts. The general applicability and appeal of the theory has been shown in the referencing by those modelling language and discourse as well as by developers of prototype systems, eg. Hirschberg(1990), Neal et al (1988), and Kobsa & Wahlster (1988a).

Some details of their theory and proposed model of dialogue will be discussed in the following sections. Given the aims of the thesis, four aspects will be considered: the workings of the theory and the assumptions it makes about the nature of communicative action and the achievement of dialogue; what phenomena are to be explained by their model and how; the source of the data and its status in the model; and the possible implementations and uses of the model in designing for human-computer interaction. First of all a brief summary of the major elements of Grosz & Sidner's theory will be presented after which the elements will be critically analysed.

2.5.1 The Theory

Theoretically Grosz & Sidner want to explain what makes a discourse - a text or dialogue - coherent and their approach stresses the role of processing and purpose (Grosz & Sidner 1986, p. 175). In order to achieve this, they argue for three components in the model: the intentional, the attentional and the linguistic. They argue that linguistic structure or structures of form beyond the sentence, which are built in the actual linear stream of discourse, must be part of a theory of discourse process. The discourse segment consisting of more than one dialogue contribution is the linguistic unit. So, linguistic structure is an important constraint on processing, but the coherence of dialogue and the unity of the discourse segment is found in the underlying intentional structure. The structure of a dialogue reflects the rational intentional behaviour of its participants. But not all intentions are discourse intentions or intended to be recognised. Grosz & Sidner are careful to restrict their notion of intention to a generalisation of Grice's utterance-level communicative intent to the discourse level. So, the initiator of a discourse segment has an intention, and the recipient must recognise that intention, through the course of dialogue during the segment, in order to have its intended effect. But linguistic and intentional components are not enough. The third component of the theory introduces the abstract concept of cognitive attentional focus, which mediates between the linguistic and intentional structures. Dialogue is the result of the recognition and coordination of intentions by more than one party in which representations of salient properties, objects and relations come in and out of focus. In this way, interpretation and production can be constrained through the attentional state mechanism. To summarise:

> *Discourse processing requires recognising how the utterances of the discourse aggregate into segments, recognising the intentions expressed in the discourse*

and the relationship among intentions, and tracking the discourse through the operation of the mechanisms associated with the attentional state. (Grosz & Sidner 1986, p. 175)

The key to Grosz & Sidner's theory is that the recognition and coordination of discourse purpose or intention by parties in the dialogue must be a central concern of any theory of dialogue process. A consequence of this is that linguistic form is relegated to signalling segment boundaries and expressing intentions. Let us look more closely at the components of their theory.

2.5.1.1 Purpose and Action

A fundamental claim of Grosz & Sidner's theory is that "intentions play a primary role in explaining discourse structure, defining discourse coherence, and providing a coherent conceptualisation of the term 'discourse' itself." (1986, p. 175) Thus, on a simple level, they reiterate the common claim that language is embedded in action and is purposeful, with the added claim that the coherence of action and dialogue is the result of underlying coordination and recognition of mental predicates like intention, plan, belief, want, etc.. The theory of discourse purpose is intimately related to extensions to Grice's notion of non-natural meaning. They posit that in dialogue participants are continually engaged in determining what each other meant by uttering a discourse contribution. The idea behind Grice's (1957) notion is to clarify the differences between the incidental transfer of information - 'given off' - and communication that is openly intended. The difference between 'natural meaning' and 'non-natural meaning' is defined[2] as:

S meant non-naturally Z by uttering U if and only if:

(i) S intended U to cause some effect Z in recipient H

(ii) S intended (i) to be achieved simply by H recognising that intention (i).

Thus, this philosophical notion of communicative intention is a special case of the more general notion of intention. For example, "a compliment achieves its intended effect only if the intention to compliment is recognised; in contrast a scream of 'boo' achieves its intended effect without the hearer having to recognise the speaker's intention." (p. 178) There are other intentions in dialogue that are not intended to communicate, for example, if a speaker intends to con a hearer then the recognition of this intention will defeat the con-trick. A summary of the basic points of Grosz & Sidner's intentional component is given below.

2 More sophisticated definitions have been developed from this in order to eliminate counter-examples.

1. The primary 'intention' is the purpose or task behind the discourse (discourse purpose=DP) or the discourse segment (discourse segment purpose=DSP). Each discourse or discourse segment has one primary 'intention' behind it.

2. The 'intention' is intended by the Initiating Conversational Participant (ICP) to be recognised by the Other CP (OCP) over a 'discourse segment', and must be recognised to be satisfied. Language is a means for achieving this.

3. The range of intentions is open-ended, but they hypothesize that structural relations between them are a small closed set. The two they focus on are: the dominance relation (DOM), or whatever satisfies one intention may partially satisfy another, and therefore this is a constituency relation; and the satisfaction precedence relation (SP), in which one intention must be satisfied before another, and this is a sequencing or ordering relation.

4. The intentional structure is distinct from the plan or task structure. It is not pre-built, but arises in the process of discourse construction.

5. The theory is derived from a generalisation of Gricean utterance-level intention theory, though Grosz & Sidner point out that they are only concerned with discourse structure and not discourse meaning. Their main concern is "with the role of DP/DSPs in determining discourse structure and in specifying how these intentions can be recognised by an OCP." (1986, p. 199)

Examples of the type of intentions that Grosz & Sidner take to be typical are in shorthand English rather than a formal language and "are intended to be a gloss for a formal statement of the actual intentions." (p. 184). Eg.

Intend A (Intend E (Tell E A (Location other setscrew))) or
Intend E (Know-How-To A (Use A wheelpuller)).

In relation to intentions, the main focus in the theoretical discussion is on beliefs and actions.

For Grosz & Sidner, the mutual intelligibility of talk and what is done through talk is a matter of the recognition by the OCP of the ICP's discourse-level intention, and recognition of the structural relations between the DSPs arising in the discourse. In addition, the relation of the discourse purposes to the production or interpretation of dialogue contributions is expressed as follows: "Our basic viewpoint is that a conversational participant needed to recognise the discourse segment purposes and the dominance relationships between them in order to process subsequent utterances of the discourse." (1987, p. 3) Communication is the recognition and coordination of discourse plans that are intended to be recognised (1987, p. 5). Ultimately, they hypothesize:

> *a discourse is coherent only when its discourse purpose is shared by all the participants and when each utterance contributes to achieving this purpose, either directly or indirectly, by contributing to the satisfaction of a discourse segment purpose.* (1986, p. 202)

The issues posed by taking such a stance are then: how is an intention expressed, how is it recognised by the intended recipient, and what conventional or shared knowledge is available for achieving the recognition and coordination? They argue that this 'process' involves linguistic cues, and propositional content and utterance-level intention recognition. Also, they need to specify the surface realisations of actions and the relations between the utterance-action and intention so that intents can be realised as actions or recognised from observable actions. Of these they have very little to say, but propose that:

> *A definitive statement characterising primary and subsidiary intentions for task-oriented dialogues awaits further research not only in discourse theory, but also in the theory of intentions and actions. In particular, a clearer statement of the interactions among the intentions of the various participants (with respect to both linguistic and nonlinguistic actions) awaits the formulation of a better theory of cooperation and multiagent activity.* (1986, p. 202)

It shall be argued later, as conversation analysts have argued, that what is required instead is a painstaking empirical analysis of the practice of dialogue. In later work, Grosz & Sidner (1987) do try and elaborate a better theory of action and intent using the notion of 'shared plan'. Rather than seeing the coherence of action as the coordination and recognition of intents as parts of individual plans, plans are constructed as shared objects separate from an individual's private plans or intents. Also, they do not determine action but provide a means of interpretation. This notion will be considered later when discussing the concept of attentional state because of the similarities in the postulation of a shared representation between individuals. It may be a useful step forward but it requires a specification of what generates action if not plans.

2.5.1.2 Activity, Structure and Process

The prime locus of dialogue coherence in Grosz & Sidner's theory is the coordination and recognition of discourse purposes. In the following, other features of the theory and how they combine into a suitable processing model are described.

It has been shown by Grosz (1977) in her work on reference that activity can in particular ways constrain linguistic interpretation and she suggests a possible computational mechanism for doing this, eg. knowledge of the task resolves 'it' and 'that' anaphora ambiguities. The earlier work is generalised from the observation that the notion of task structure is essential for understanding task-oriented dialogues to a basic assumption of the theory: that the dialogue participants regularly engage in purposeful activities that structure

the dialogue in analogous ways to task dialogues. These activities can be discourse specific, physical or mental. However, the task and the dialogue activity are distinct; the structure of discourse segment purposes is a result of the dialogue and may have a different structure from the task plan as well as a different level of detail. This generalisation is then filled out by relating the coherence of activities to the mentalist account of purposeful action outlined earlier: thus, not only are task-based activities purposeful and intentional, but so are activities that are intrinsic to the dialogue itself.

The linguistic component is "the structure of the actual sequences of utterances in the discourse" made up of discourse segments or chunks of 'utterances', viz. "the actual saying or writing of particular sequences of phrases or clauses." (1986, p. 177) It is a segmental categorisation of discourse leading to a hierarchical organisation of discrete segments. Discourse is broken down into segments in which utterances function in the segment to further its discourse purpose. The motivation for connecting structure with purpose is "What individuates a discourse? What makes it coherent? That is, faced with a sequence of utterances, how does one know whether they constitute a single discourse, several (perhaps interleaved) discourses, or none?" (1986, p. 175) It is unclear if they mean for the analyst or the participants.

Finally, the attentional component is an "abstraction of the focus of attention of the discourse participants", which mediates between the intentional and linguistic levels. They claim that "attention is essential in explicating the processing of utterances in discourse." (1986, p. 175) It provides a central locus of control by which discourse segmentation is achieved as the product of recognising and coordinating discourse purposes. It also provides a constraint on processing by making clear what elements are currently relevant. A stack model is used to process contributions and the mechanism provides for a hierarchical organisation of discourse segmentation. Contributions can push salient discourse segments and their features onto the stack or pop them off to return to held-over segments requiring completion. The attentional state is "a property of the discourse itself, not of the discourse participants" (1986, p. 179). So, it is a postulated shared structure of discourse that is theoretically separate from the mental states of an individual and emerges in the dialogue. This is an important distinction that has since become topical in AI models of natural language, ie. the dialogue record or score-board that is shared is separated from the 'individual' user model, eg. private intentions and beliefs concerning others. Basically, the claim is that participants if they are to engage in mutual activity must build shared representations that are publicly inspectable. The focus mechanism and the notion of 'shared plan' are two such representations that maintain a record of what beliefs are mutually known and what objects are salient at a particular point in the dialogue. In the model these are computational

representations that are maintained by each participant on the assumption that they are mutually accessible; in fact, they are internal and not externally available nor authorised as is the case in sports games for example. This is an interesting development but unfortunately merely hides the problem of how mutual knowledge is achieved and maintained behind a metaphor based on mutual access. Even if operations on the attentional state are identical for cognitive individuals it is not true that the 'shared state' of each individual will correspond in practice.

Grosz & Sidner's theory is trying to be very comprehensive in coverage. There are several clear improvements on other computational models, like plan-based speech act theory and rhetorical theories. For instance, they recognise a distinction between the intentional structure and the plan structure. Previous accounts tended to conflate the intentional with the task or plan structure and thus could not distinguish between what one knows about the task and one's intentions for performing it. Thus the attentional component focus model was redundant as a 'mediator' between the intentional structure and the task structure, and could only serve as the 'realisation' interface, ie. the task plan was the same as the discourse plan. In these accounts the emergence of purposes related to the task but specific to the dialogue itself, eg. contingencies in the coordination of communication about the task, were unaccounted for by the model. They also note the contingency of discourse structure and its emergence during the discourse as a property of the discourse and not the participants, even though the theoretical direction taken to explain this is problematic.

2.5.2 What Can Grosz and Sidner's Model Explain?

Grosz & Sidner apply their theory to two types of dialogue phenomena to illustrate the type of explanation possible by having a three component model of discourse structure. Unfortunately, they do not get at the richness of the phenomena that other disciplines have uncovered.

2.5.2.1 The Structure of Interruptions

One dialogue phenomenon they look at commonly comes under the category of 'interruption'. Their 'definition' of an interruption has two forms of which the strongest is: "An interruption is a discourse segment whose DSP is not dominated nor satisfaction-preceded by the DSP of any preceding segment." (1986, p. 192) They give two examples of this type. First, the 'true interruption' in which a discourse segment purpose (DSP) is pushed onto the focus stack. No relation of dominance (DOM) or satisfaction-precedence (SP) exists between the DSP and lower members of the stack, and access to entities in the lower members is not allowed in this case. Second, the 'digression' which is as above except there is a common focus or

salience, eg. an object referred to in both segments, between the focused DSP and the prior DSP[3].

The expectation of a return to the interrupted segment DSP is built into the stack model, but it becomes more of an effect of the stack model than a normative expectation that is oriented to by the participants and hence reproduced. It is on this point that the model fails to provide an adequate account. The sense in which interruptions are produced and managed by all participants in the talk is lost to a mechanical model that takes the environment of an interruption to be a prior DSP. The interruption arises simply because there is no structural relation between the new DSP and the prior. A limited number of conversational cues are available to signal its absence.

2.5.2.2 The Function of Clue Words and Phrases

Because of the design of the model, 'cue phrases', eg. 'anyway', 'however', serve a certain function in discourse structure. They are an abbreviated, indirect means of indicating change in the attentional state and intentional structure. As discourse evolves through 'sequence time', which is modelled by the stack mechanism, cue phrases can indicate sequencing or constituency relations between DSPs, as well as an attentional state change. Also, changes to the attentional state that do not involve intentional structure changes are possible. The cue phrases are employed by the initiating conversational participant (ICP) to indicate discourse segment boundaries to the other conversational participant (OCP). For example, if a stack currently exists with a DSP A in focus, then what intentional structure changes can cues enable coincident with a new DSP B being pushed onto the stack?

a) A dominates B (constituency relation), eg. "for example".
b) no relation (true interruption), eg. "excuse me".
c) B satisfaction-precedes A (sequencing relation: flashback), eg. "i forgot about".
d) no relation but common focus (digression), eg. "by the way".

Note that b) - d) all relate to interruptions and it is claimed they are not rhetorical. Cue phrases can also indicate attentional state change only.

3 This is an unusual property of a stack because it is a relaxation of the defining properties of a stack that normally prohibits access to lower members.

e) push DSP onto stack, eg. "now", "and", "but"

f) pop DSP(s) from stack, eg. "anyway"

g) complete a DSP, eg. "ok"

2.5.3 Dialogue Examples as Data

Grosz & Sidner have tried to connect their theory with records of discourse. However, it is unclear whether these are the prime motivation for developing the theory or simply to illustrate its explanatory power. The main examples in the paper - a text and a dialogue from a corpus - are not intuitive, but others are, eg. the interruption example (1986, p. 192). The dialogue example can be seen in Figure 2-2 and is used to illustrate their analysis in terms of discourse segment decomposition through the process of intention ascription and attentional focus change. Later in this chapter an account is given of the problems encountered when an alternative conversation analytic analysis was attempted.

==========

FIGURE 2-2 - Grosz & Sidner's data example

```
(1)  E: First you have to remove the flywheel.
(2)  A: How do I remove the flywheel?
(3)  E: First, loosen the two allen head setscrews
        holding it to the shaft, then pull it off.
(4)  A: OK.
(5)     I can find one screw. Where's the other one?
(6)  E: On the hub of the flywheel.
(7)  A: That's the one I found.  Where's the other one?
(8)  E: About ninety degrees around the hub from the first
        one.
(9)  A: I don't understand. I can only find one.  Oh
        wait, yes I think I was on the wrong wheel.
(10) E: Show me what you are doing.
(11) A: I was on the wrong wheel and I can find
        them both now.
(12)    The tool I have is awkward. Is there another
        tool that I could use instead?
(13) E: Show me the tool you are using.
(14) A: OK.
(15) E: Are you sure you are using the right size
        key?
(16) A: I'll try some others.
(17)    I found an angle I can get at it.
(18)    The two screws are loose, but I'm having trouble
        getting the wheel off.
(19) E: Use the wheelpuller. Do you know how to use
        it?
(20) A: No.
(21) E: Do you know what it looks like?
```

```
(22) A: Yes.
(23) E: Show it to me please.
(24) A: OK.
(25) E: Good. Loosen the screw in the center and
        place the jaws around the hub of the
        wheel, then tighten the screw onto the
        center of the shaft.  The wheel should
        slide off.
==================================================
```

2.5.4 Implementation of the Model

> *We did not propose an actual processing model. A computational theory of the recognition of discourse segment purposes depends on underlying theories of intention, action and plans.* (1987, p. 3-4)

With respect to implementation issues, Grosz & Sidner claim that they are only concerned with an abstract model of discourse structure. Connecting theory with linguistic evidence through either constructing a dialogue system or a psychological theory of language use is a possibility, but not entertained. Grosz & Sidner hope that their abstract theory will be implementable and thus be used in simulating human conduct. However, they do consider some of the issues they must confront if the theory was used to build a computational model. To do this they claim that it "requires determining how each of the individual components projects onto the model of an individual discourse participant." (1986, p. 188)

2.5.5 An Example of the Model in Action

Consider the example quoted above in Figure 2-2. It is rather complex but the workings of parts of the model will be illustrated based on the segmentation proposed by Grosz & Sidner (1986). How does the segmentation into primary discourse segments get achieved in the dialogue itself? For example, the whole fragment is ascribed as one discourse segment DS1 initiated by the apprentice A. Its primary DSP is: the expert E (ICP) intends that the apprentice intend to remove the flywheel. It is claimed that this is expressed directly in the utterance-intention on line 1. Also, the cue phrase 'first' marks the DSP as the first of several intentions whose satisfaction will contribute to satisfying the larger discourse. This segment is pushed onto the attentional state focus stack with the flywheel in focus. Lines 2-25 are then coordinating the recognition by the apprentice A of this discourse intention which happens to include several subordinate segments. In line 5, A initiates another discourse segment DS2 with the DSP: A intends E to intend to tell A the location of the other setscrew. This is also meant to be clear from the utterance-level intention of line 5. This discourse segment is pushed onto the stack with the screw now in focus. A relation of dominance is posited between the segments DS1 and DS2 on the stack because of the shared knowledge

of the two participants and the commitment that A has made to removing the flywheel. By classifying utterances in terms of their contribution to discourse purpose recognition and dialogue segmentation, Grosz & Sidner explain apparent ambiguities. For example, in line 25 the phrase "the screw in the center" is potentially ambiguous between referring to one of the screws on the flywheel in line 18 or to the screw on the wheelpuller. However, it is claimed that the segmentation posited by the model and the current focus on segment DS5 (lines 19-25) makes the reference of the phrase clear.

2.6 Problems and Limitations of Grosz & Sidner's Theory

Grosz & Sidner attempt a comprehensive theory of dialogue, but there are a number of problems and limitations. Some of these will be focused on in the remainder of this chapter. Their work is particularly problematic as an adequate theory of human dialogue and also as an appropriate model for the design and conception of dialogue artifacts. In the following, the weak theory of action and activity will be analysed in detail first. The phenomena explained by Grosz & Sidner's model will be re-examined in a different light. Then the weaknesses in the empirical methodology characteristic of computer theories of dialogue will be illustrated and discussed, and the possibilities for implementation and artifact design considered. Finally, a reconstruction of their theory in an interactional vocabulary will show that their account is founded on a rationalised mentalist account of dialogue activity in which the situated achievement of that activity is unaccounted for.

2.6.1 Theoretical Problems

Problems begin with the intentional component that underlies the coherence of dialogue. An intentional theory of rational, purposeful action forms the backbone of the model. Because of the recognition that linguistic models do not suffice, earlier accounts of dialogue have tried to find formal systems of speech action that could explain the coherence of dialogue, but they suffer from many problems. Levinson has criticised speech act models and discourse analysis in general, arguing that they lead to ad hoc theorising and a lack of empiricism (Levinson 1981a and b, 1983). Contemporary process models are more sophisticated but still need careful examination. Suchman (1987) and Agre & Chapman (1989) provide extended critiques of the conception of action and plans in artificial intelligence and cognitive science and the thesis is concerned with continuing their criticisms with respect to computational dialogue models, and to suggest some missing details necessary for an adequate account.

The most prevalent model of action in cognitive science and artificial intelligence has a long heritage in traditional rationalist Western philosophy and science, as Winograd & Flores

(1986) have pointed out. Grosz & Sidner (1987) fall back on the plan-based model in which dialogue coherence is to be explained in terms of the recognition and coordination of underlying plans of actions. However, is discourse intention recognition and coordination the basis for discourse structure and coherence or the basis by which participants construct or 'process' discourse? It is true that analysts and participants do construct intentional explanations in retrospect, but that does not mean it is a pervasive component of taking part in that conduct. Criticisms of the plan-based model of action are to be found in Suchman (1987) who suggests that plans are powerful for reconstruction and reflection, but necessarily poor for the actual doing of action, and in Wynn (1980, p. 88):

> *Purposes can be, and typically are discovered in the course of interaction rather than planned. Purposes are thus emergent from interaction rather than a priori organising principles of it.*

Following the line of Suchman (1987), I will attack three basic assumptions that underlie Grosz & Sidner's theory that is claimed to explain the mutual intelligibility of action and consequently dialogue structure. These are:

1) The significance of action is derived from the mutual recognition and coordination of intentions - this handles the relationship of observable behaviour to intent.
2) Conventions for the expression of intent - some form-intent correlation - enable the recognition by the OCP of the ICP's intention.
3) The stability of meaning is achieved by a shared background knowledge.

The **first assumption** is primary for any plan-based account of agency and cooperation in AI. Intuitively, it seems to be a sensible statement given that we often use the words 'believe', 'want', 'act', 'intend', etc. in talk and discourse. Our intentional vocabulary of plans, goals and beliefs is rich and subtle for describing, talking about and thus reflecting on the purposeful behaviour of others and ourselves. It is very helpful in projecting courses of action, but ultimately vague in terms of the actual 'doing' of action. A plan is a powerful abstraction because the indefinitely detailed situated actions do not need to be expressed; if such expression were possible. Not in the sense that subplans need to be formulated to deal with problems in the course of action, but that plan or sub-plan are necessarily indeterminate in specifying the details of actions, ie. the actions are underspecified. They are also interesting when used in a post-rationalisation of action's circumstances. For example, the court-room is the setting for constructing 'what happened' from the accounts of witnesses.

> *Our imagined projections and our retrospective reconstructions are the principal means by which we catch hold of situated action and reason about it, while situated action itself, in contrast, is essentially transparent to us as actors.* (Suchman 1987, p. 39)

The danger lies in the reification of our descriptions into the place of action's actual circumstances, and taking the explicit AI procedures as an explanation of the significance of those situated actions. Grosz & Sidner's analysis of the task dialogue does give the impression of a post hoc rationalisation that can be continued ad infinitum. This phenomenon, also known as 'indefinite glossing', is easily illustrated because it is an unavoidable practical activity engaged in routinely by everybody (Garfinkel 1967). Grosz & Sidner focus on the 'primary' intentions behind the discourse segments. How were these decided? Can they be expanded upon? How many 'secondary' intentions are there? Given the following quotation, how much filling in of the subtasks must the OCP have to do?

> *...in adopting the intention to carry out that action, the OCP also intends to perform whatever subactions are necessary. Thus once the apprentice intends to remove the flywheel, he also commits himself to the collateral intentions of loosening the setscrews and pulling the wheel off. Note, however, that not all the subactions need to be introduced explicitly into the discourse. The apprentice may do several actions that are never mentioned, and the expert may assume that these are being undertaken on the basis of other information that the apprentice obtains.* (Grosz & Sidner 1986, p. 185)

Not only could the analysts indefinitely gloss the 'underlying' intentions behind talk and action in the task dialogue, but also the participants could expand in a similar fashion should some trouble with the action's course or circumstances arise, or should he feel like it. What is interesting is when and why do the participants invoke an intentional vocabulary? Does it occur only when the transparency of talk suddenly becomes opaque? What are the interactional consequences of the intentional talk? For example, in Lynch (1985, p. 210-212) the use of an intentional vocabulary in modifications of accounts of 'objects' during disagreement sequences for displaying agency or uncertainty is illustrated.

The claim here is that the first assumption underlies no explanation of the achievement of talk or action in situ[4]. In fact, much talk seems to be unanalysable with this model. What account can Grosz & Sidner give in talk occurring in doctor-patient, counsellor-interviewee, court-room settings in which the notion of 'agenda', 'plan', or 'intention' is commonly not recognised by the OCP? In these settings, Suchman (1987) has argued that the agenda is not shared in the sense of it being recognised and coordinated. One party to the talk is largely unaware of the other's reason(s) or intention(s) 'behind' the interaction. That party is responsive to the other locally, and interactively constitutes the agenda and discourse

4 Indeed Du Bois (1987) argues that some meaningful language activities in some cultures can only be understood if they are without intent.

structure without knowledge of such a structure. The structure is the product of the routine, transparent, local, orderly work of the participants.

An attempt is made in Grosz & Sidner (1987) to expand on the observation made in the earlier paper about the shared nature of discourse by positing the notion of shared plans constructed in collaboration. This is more subtle than most plan-based accounts of action in that plans are supposedly not individualistic and do not determine actions but restrict the interpretation and production of action. This is to be commended but still suffers from a heavy emphasis that a shared plan is necessary to engage in mutual action and that intention recognition does underlie the dialogue activities participants are engaged in. There is no alternative in the model.

> *Discourses may exhibit two types of collaborative behaviour: collaboration in the domain of the discourse and collaboration with respect to the discourse itself... surface collaborations (eg. coordinating turns in a dialogue) but also collaborations related to discourse purpose.* (1987, p. 4)

This is a positive statement. However, they do not seem to acknowledge the importance of the 'surface' collaboration for the constitution of action itself. This is important because,

> *rather than depend on the reliable recognition of intent, mutual intelligibility turns on the availability of communicative resources to detect, remedy, and at times even exploit the inevitable uncertainties of action's significance.* (Suchman 1987, p. 69)

The **second assumption** has again a long lineage in the rationalist tradition of theories of action and language. Given the first assumption, and consequently the recognition of intentions as a real and continuous problem, some way must be found to connect up the language in expression, or observable behaviour as action, with the underlying intentions. A typical solution is to assume a 'conventional' procedure for a bijective mapping from one to the other. Levinson (1981b) gives an in-depth critique of any attempt to relate form and function in this way.

Grosz & Sidner base their theory on a Gricean account of utterance-level intention that requires, for Grosz & Sidner, a specification of how any utterance can come to have a particular intention. Unfortunately, such a formal account is not forthcoming, as argued by Levinson (1981b, 1983), at the utterance-level so it is hard to see how "the modes of correlation between features of the discourse segments and types of discourse-level intentions" (Grosz & Sidner 1986, p. 199) can be specified in advance. Within the framework the proposal for more investigation into the formal mappings from surface forms to a deeper level of intentions is missing the point. In fact, it has been shown in conversation analysis that interpretations of utterances can only be found provisionally and locally on each occasion of production.

It is interesting to wonder how the intention categorisation and discourse segmentation that is specified in Grosz & Sidner (1986) for specific examples could be justified. The ascription of intentions and structure implicitly draws upon intuition: "the researcher must use his or her knowledge of the world as a resource for interpreting discourse and textual materials" (Cicourel 1980, p. 127). However, intuition should need empiricism - the reality of a structure, a segment, is not considered empirically by Grosz & Sidner - and a theory that borrows from Grice's utterance-intention notion is fraught with the difficulties of linguistic form and utterance force correlation. Another crucial problem, that is recognised by Grosz & Sidner, is the pre-ordained nature of segmental intention recognition. That is, only one primary discourse purpose is to be recognised for each discourse segment after its initiation. For example in the dialogue example in Figure 2-2, lines 18 to 25 are part of a discourse segment with the discourse purpose that A, the ICP, intends that E, the OCP, intends to tell A how to get the wheel off. During its life on the attentional stack a discourse segment and its purpose are in the process of being communicated and recognised by its recipient. Thus the receiving conversational participant has a conflict between not knowing what the purpose is until a segment is complete and knowing enough in order to complete it. This could be explained as the consequence of over-rationalising and reifying the vocabulary of post-hoc explanation into a generatory mechanism. Hence, over-rationalisation of dialogue activity and the focus on an explicit individual process leads theory to an impasse.

The **third assumption** is that there exists a background of shared knowledge implied and presupposed in action that gives it significance and stability of meaning. Grosz & Sidner's claims on this front are an implicit part of their whole approach, and are innocent enough at first glance. It is argued that

> *the apprentice and expert share certain knowledge about the task. Some of this shared task knowledge comes from the discourse per se ..., but some of it comes from general knowledge, perceptual information, and the like.* (1986, p. 186-7)

Suchman (1987, p. 35) summarises this viewpoint as,

> *... the image evoked by 'shared knowledge' is a potentially enumerable body of assumptions or presuppositions, that stands behind every explicit action or utterance, and from which participants in interaction selectively draw in understanding each other's actions.*

However, Garfinkel (1967) has shown in his famous experiment on glossing, that indefinite accounts of actions or talk can be given. But, such a practice is not relevant to participants in talk in situ, unless troubles occur and accounts become relevant, whereupon they can be constructed. In the experiment subjects were asked to explicitly describe what presuppositions or knowledge the participants in a transcribed dialogue would possibly need to have used in order to explain why the dialogue turned out as it did. Subjects on completing

the task were asked to fill in more detail, again and again; they finally complained that there was no end to this procedure. A contrary approach, ethnomethodology, concentrates on the question of how the participants make 'sense' in the local circumstances on each occasion, not what theoretical social mechanisms 'underlie' its achievement. Many accounts in artificial intelligence and cognitive science try to over-rationalise in exactly the latter manner in order to locate explicit 'cognitive' procedures that by their very nature require this enumeration[5]. Thus they are susceptible to the

> *indefinitely extendable and unverifiable categorisation and speculation about actor's intents so typical of DA-style analysis. [DA= Discourse analysis]* (Levinson 1983, p. 319)

Let us work through some important points and consequences of the criticisms of the three assumptions above. Everyday 'rational' conduct itself, like dialogue, is routinely done without explicit intents and their recognition.

> *An analytic focus on direct experience in the lived-in-world leads to emphasis on a reflexive view of the constitution of goals in activity and to the proposition that goals are constructed, often in verbal interpretation.... This retrospective and reflexive character is not compatible with a linear view of action as directed towards established goals. It suggests that action is not 'goal directed' nor are goals a condition for action.* (Lave 1988, p. 183)

For example, communicative intent usually becomes explicit in dealing with 'trouble' where the flow of conduct is disrupted and practical reasoning about intents and plans, etc. can usefully come into play in the repair of 'what was meant or intended'. The intentional vocabulary is then available for descriptions, formulations, 'post hoc' rationalisations, etc..

> *Plans, in other words, are just an efficient way of projecting, interpreting and classifying behaviour, without taking any strong stand on underlying cognitive mechanism. The point of intentional description, in this folk psychological sense, is not to control actions, but to comprehend them.* (Suchman 1986, p. 5)

But this is not to deny purposeful activity and the relevance of intentions. Giddens (1979) looks at the senses in which meaning is used in ordinary English usage: what an actor means to say or do, and what the meaning of his utterance or act is. He notes the tendency for a reduction to take place. Meaning is either what they mean or intend to say, or what speakers mean to say is irrelevant to an understanding of the nature of meaning.

---

5 It should be noted that there are interesting parallels between the conceptual intentional theories of communicative action and the conventional signal theories of turn-taking. Both need notions of convention, recognition and expression criteria, and explicit acts. See Wilson et al (1984) for an analysis of several models of turn-taking in conversation.

> *I regard the meanings of communicative acts - that is, acts in which one element of the reflexive monitoring of conduct includes the intent to communicate with another - as in principle distinguishable from, and are sustained in, the differences expressed in the practices of language-games; but such practices, as the active accomplishment of human subjects, are organised through and in the reflexive monitoring of conduct. The interplay of meaning as communicative intent, and meaning as difference, represents the duality of structure in the production of meaning.* (Giddens 1979, p. 85)

Crucially, "intentions are only constituted within the reflexive monitoring of action, which however in turn only operates in conjunction with unacknowledged conditions and outcomes of action." (Giddens 1979, p. 42) Thus, we need to study how plans and representations are brought into productive interaction with the contingent unrepresented circumstances of situated action, given that action does not just occur in specific settings or circumstances; it is constituted within them and reproduces them as contextual relevances.

Also, it is important to note the conceptual and representational bias in computational theories. That is,

> *For centuries epistemologists have concentrated primarily on conceptual representation and its problematic relation to objects in the world, assuming that representation is cognitively prior to all else.... An epistemology that begins with activity and perception, which are first and foremost embedded in the world, may simply by-pass the classical problem of reference — of mediating conceptual representations. Conceptual representation can be seen as secondary, growing out of the direct 'negotiations' people conduct with the physical and social world to which they have direct access in activity.* (Brown et al 1988, p. 32-33)

It has been argued by some researchers in artificial intelligence and computational linguistics that all three components of Grosz & Sidner's model should be represented, but without necessarily adopting their particular attentional process mechanism. The main point is that as much as possible that can be conceptualised of dialogue structure should be represented. For example, it is reported in Kobsa & Wahlster (1988a, p. 88) that some researchers have claimed that a discourse model should contain "everything that should be derived from an analysis of discourse, to present a representation for the structure of the discourse, useful in subsequently responding." Dialogue components are some of the latest to be included on the collective score-board of shared 'knowledge' that was criticised earlier. The structures represented on the 'shared score-board' of dialogue are restrictions on the possible interpretations. The aim is to include all the possible representations of structure and 'knowledge', in order that a participant in a period of reflection can decide what to say and how to say it. In Grosz & Sidner, the attentional state is the mediating score-board. This is another example of over-rationalising the activities we engage in. (See Hahn (1986) for a recommendation for close connections between analysis, representations and evaluation.)

Structural models may describe or conceptualise phenomena that are not available to those engaged in the activity and this may of course be useful for a model that is interested in the production and interpretation of conduct for pragmatic reasons, like building an artifact that must accomplish robust behaviour, but it needs to be shown that it is useful, viable, tractable and pertinent to include it.

In addition to the problems encountered above, the model, by focusing on conceptualisations and pre-design, is not in the position of acknowledging or including emergent properties of social interaction. For example, Cicourel (1980) has argued that,

> *the problem-solving or schema model is preoccupied with the individual's knowledge base and the explicit attribution of goals, plans, intentions, actions, and motives. The speech act categories and process descriptions of actions can be used to create complex predicates that address the participants' use of knowledge in pursuing goals, plans, and actions, and the intentions and motives that seem to orient these objectives and activities. But the model does not identify emergent properties of discourse or interaction that are part of the local production of everyday life.* (p. 128)

Grosz & Sidner do consider dialogue as an activity, which derives from an earlier interest in the effect of task activities on dialogue structure, but they do not have a well developed theory of activity. Generalising from a philosophical theory of meaning they try to found an intentional theory of dialogue activity, but this means that the local dynamics of dialogue that are not predictable and emerge in the engagement in dialogue itself are unaccounted for in the theory and the model[6]. It is the focus on a deeper level of explanation in computer models of dialogue coherence that reproduces the 'conduit metaphor': our everyday vocabulary for discussing language communication tends to distort any analytic explanation of that conduct. The metaphor unfortunately takes linguistic communication as the transfer of linguistically wrapped mental packages between a producer and a receiver that requires a shared linguistic code to encode and unwrap the message (Reddy 1979). Thus local and emergent features of the contingent achievement of dialogue in situ go missing. In a way this parallels the distinction between competence and performance in Chomskyan linguistics. Intention recognition and coordination is the meat; the realisation in language and interaction is the gravy. The 'conduit metaphor' and the 'interchange model' in computational terms tend to view talk as intentional speech exchange involving linguistic, symbolic communication by rational cognitive processors. The computational theory of communicative action has

6 The concept of dynamics is illustrated by Simon (1970) with the example of the emergent complex behaviour of an ant in its local ecology; also Agre (1988) discusses the importance of a situated theory of dynamic activity. An example ilustrating the dynamics of dialogue will be taken up in Chapter 4 and Chapter 5.

consistently taken-for-granted situated interaction because of the conceptual formulations of dialogue, action and discourse. This problematic development may have occurred partly because most computer models of dialogue have their origin in the study of distanced written textual discourses - in which the debate about the role of the writer's intentions in production and interpretation when the reader is in a different time-space context is prominent and also the early influence of communication theory on linguistics.

The attentional component is a traditional attempt at using the process-like notion of experiential 'conscious focus' in a computer model and the linguistic component is almost a mere formal realisation of the intentional structure. There are hardly any dynamic ties between language organisation and dialogue structure. For example, contributions are serving only one function in the model and that is to complete the recognition of the discourse segment purpose. The attentional state does result in a separation of the task plans from the emergent intentional structure achieved in dialogue, but it seems too restrictive to allow only the hierarchical embedding of complete discourse segments (Bateman, personal communication). In addition, Fox (1987) argues that the notion of 'interactional reconstruction' must be accommodated in models of dialogue. She draws from conversation analytic studies to point out that analyses or assignments of structure should not be unchangeable. In everyday talk it has been shown that contributions are not assigned a structure or meaning that is communicated, but that interpretations of the relevance of a contribution are negotiated and renegotiated in next and later talk. Thus if a model of dialogue assigns a fixed structure on the basis of one interpretation - for example, the assignment of a contribution to fulfilling a particular discourse segment purpose - how can it accommodate modifications that become necessary in later dialogue? This problem arises of course because Grosz & Sidner's model gives priority to the underlying purpose of a discourse segment that is not found or emergent but is to be communicated - first comes the intent and then the realisation.

This section has put forward many problems and limitations with the theoretical side of Grosz & Sidner (1986, 1987), which unfortunately is far too general for computational consideration. Even so, the bulky and complex structural baggage is still too static to accommodate Fox's notion of interactional reconstruction (Fox 1987) or Anderson & Garrod's explication of the dynamics of negotiated meaning, that require a model to have local dynamic structures (Anderson & Garrod 1987). The concentration on purposeful interchange dialogue, speaker production and knowledge structures has led to the backgrounding of the interactional domain. Dialogue is not, however, alternating action-reaction interchange, but "joint action accomplished through the participant's continuous engagement in speaking and listening." (Suchman 1987, p. 71)

> *Coherence is not just a conversational accomplishment to be met through location, content, and the structure of an answer, but an interactional accomplishment to be met through the actions and displays performed through locally situated moves.* (Schiffrin 1987, p. 119)

Grosz & Sidner's model, like many computational models, is independent of interactive circumstances, even though the dialogue example contains particulars and structures emergent from the circumstances of participants and their joint activity, as discussed later in this chapter. It is maybe pertinent to consider instead a practical approach in which interaction, joint negotiation and 'making do' have more consideration rather than rational planning in advance.

2.6.2 The Explanation of Dialogue Phenomena

In the following a criticism of the specific details of the theory in terms of the explanation of particular phenomena will be given.

2.6.2.1 Interruptive Activity

In order to illustrate the explanatory limitations of Grosz & Sidner's model and demonstrate the local work of achieving dialogue coherence and structure, an alternative perspective will be described. Conversation analysis (CA) has examined in detail the complex, detailed achievements involved in managing interruptions. For example, Jefferson (1972) has tackled in depth the same sort of phenomena as the 'digression'described above from Grosz & Sidner (1986). Her analysis is typical of CA's empirical approach that attempts to identify intrinsic and unobvious features of talk with as little conceptual theorising as possible. In the paper, the interest is in those occurrences of talk that constitute a 'break' in the activities, followed by a return to those activities. The intuition explored in the analysis is that those occurrences are not 'part' of the activity but are somehow relevant. In contrast to Grosz & Sidner, the questions of importance are as follows. How do the participants understand and manage the 'side sequence'? Are there characteristic structural 'parts' in the whole sequence that participants orientate towards? How is the ongoing activity returned to, and what conversational work must be done by the participants to achieve this? What are the responsibilities for carrying the sequence through? Jefferson argues, from real examples of conversation, that a three part sequence is demonstrably oriented to by the participants. The first part is the "ongoing sequence" (O), the second part is the "side sequence" (S), and the last part is the "return to ongoing sequence" (R). The sequences may be exploited or not, but they are argued for as a normative expectation which analysably informs the ongoing talk. An example will illustrate.

[Example III in Jefferson (1972, p.317)]

	1	O	A: An' everybody's askin 'im t'dance.
	2	O	B: An' because he's scareda dancing he's gonna dance in private til he learns how.
	3	O	A: And a goodlooking girl comes up to you and asks you, y'know,
	4	S	B: "Gi(hh)rl asks you to-"
	5	S	B: Alright,
	6	S	C: Well it's happened a lotta ti//mes,
>	7	S	B: Okay okay go ahead.
	8		(1.0)
>	9	R	B: So he says "no."
	10		(1.0)
	11	R	B: Cause he's scared to admit that he can't dance an' he's scared to try. Cause he's gonna make a fool of himself.

At line 7, B's "okay okay go ahead" is typical of a side sequence terminator leading into the return to the ongoing sequence. In this case, the return is problematic as the silences at lines 8 and 10 show. Now, Jefferson argues that there are two particular ways in her data of returning to the ongoing activity. 'Resumption' occurs when there is an explicit marker of a problem in accomplishing a return, while 'continuation' is specifically directed at 'covering up' the problem or leaving it transparent. Continuation retrospectively suggests a reinterpretation of prior talk. In the example above, the "so he says 'no' " at line 9 is an attempted return under the guise of 'continuation' which is deleting the S and tying in with the ongoing sequence O. Thus, in the terms of Grosz & Sidner's model a segment is reclassified - what was a digression is now a continuation in the ongoing sequence. It is difficult to see how this can be explained in their model. How can a discourse segment 'digression' that is initiated with a specific intent and pushed onto the stack come to be a coherent sequence in which it is not a separate segment at all?

Similarly, Grosz & Sidner's other definition of 'weak' interruption can be investigated empirically to see what happens in talk, as can all conversational phenomena of this type that involve the management of current activity in relation to prior, that seems intuitively not to be a 'part' of that prior activity. Taking the Grosz & Sidner divisions for a moment, another set of phenomena that are open for empirical investigation is where talk is concerned with current activity that is related to prior activities in some way, maybe constituency or ordering. For example, insertion sequences, flashbacks, clarification sequences and elaboration sequences. In fact, the resources available for the situated achievement and management

of activities and their relationships are the phenomena for investigation, and any mechanism or theory should be derived from the richness of the data, not by introspection or intuition.

2.6.2.2 Discourse Markers

It is not clear in Grosz & Sidner from what source the characterisations about the function of language forms like "anyway", "however", etc. are derived from. It is never said whether they are intuitions, or deductions from records of natural talk or texts, or analyses culled from prior work that are to be usefully explained in their theory. Besides, the gross explanations of the phenomena do not do justice to the local interpretations demonstrated in conversation analytic research. Also, it seems that even though their model is distinctly better than other computational models because these forms, cue words and phrases, are functionally tied in with discourse sequencing, the mechanism of the attentional state is influencing the explanation of phenomena adversely. The stack mechanism just does not do justice to the complexity of the functioning of a marker. Consequently, the rich linguistic and interactional environments are not explored, but ignored in the hope of a theory of dialogue based on intention recognition underlying the use of markers. One might argue that Grosz & Sidner are not interested in the complexities of cue phrases as such, but only in the role of them in constraining intention recognition. However, besides the complications in attempting an abstract account on which to base discourse coherence, other approaches suggest that an attack from an empirical base with as little theoretical apparatus as possible is best when language forms like these are studied. For example, from both linguistic and conversation analytic perspectives. Levinson & Owen (1981), Schiffrin (1987), and conversation analytic work by Heritage (1984b) and Schegloff (1982) among others, give a different picture of the role of these forms in the achievement of coherent dialogue.

Some of the cue phrases explained by Grosz & Sidner are also investigated by Levinson & Owen (1981). For example, "by the way" is interpreted as a 'digression' in Grosz & Sidner - a focus space is stacked and no dominance or satisfaction-precedence relation is established between the new segment and previous stacked segments, but a common focus is maintained between the segment and the one below. Levinson & Owen argue that this marker may occur when a conversational activity is seen to be less important than a prior activity, sequentially not warranted, and that the prior activity is suspended. The design of the new activity, and the environment of the marker are very important and are detailed in each case. These interactional sensitivities are not available to Grosz & Sidner. They posit that some 'topical' commonality is maintained across the focus spaces. This is not necessarily the case, and is not important for some environments of the marker "by the way".

Also, Levinson & Owen argue for several conversational functions of the marker 'anyway' which involve the notions of ordering and presentation of activities, and the issue of conversational closure. An initial analysis proposes that 'anyway' has the following interpretations in talk:

> Anyway1 - activity resumption, eg. closing initiation with a return to the reason for a telephone call.
> Anyway2 - offer to close.
> Anyway3 - misplaced activities, that should have occurred, but treated as having occurred and now resumed.

Grosz & Sidner might claim a similarity between anyway1 and their explanation - that 'anyway' is a cue functioning to pop a segment off the attentional stack - but they cannot detect the subtle differences between the three above that their model would equivocate. For instance, anyway3 shows how markers can be exploited for conversational and interactional reasons, but Grosz & Sidner's model can say nothing about a pop to a segment that never existed as happens with anyway3.

2.6.3 A Weak Empirical Methodology

Having a dialogue fragment is crucial in testing and illustrating a theory or claim, so a non-intuitive example is to be commended. However, difficulties were encountered when trying to analyse Grosz & Sidner's dialogue example according to their model, and when trying to re-analyse it in more familiar terms. It seemed unlike human dialogue that occurs in the routine circumstances of everyday life. Expectations lead to a feeling of flatness: the lack of interruption or overlap, the odd responses, and omission of time or non-linguistic action from the script.

A re-analysis of the data fragment that Grosz & Sidner (1986) use and which is the main dialogue example for their paper was attempted. Unfortunately, it is problematic because the setting of the example is not what at first it appears to be given the context of its presentation. It was first thought that the transcript was of two participants who were in the same room involved in a practical activity, but this hypothesis was soon dismissed because the dialogue did notaccord with that view. For example, lines 10, 13, and 23 in Figure 2-2 (Section 2.5.3) are rather odd if they did occur in a dialogue in the expected setting. It was then conjectured that the example must be a transcript of an activity in which communication

was conducted over the telephone with some limited visual contact through a window. Several analyses from different perspectives were attempted. But it would appear from studying Deutsch[7] (1974) that some experiments were done at SRI laboratory in the 1970s in the context of the design of a computer 'expert consultant' to a human apprentice. The experiments were designed in order to derive human protocols similar to those envisaged in the human-computer environment. A task was chosen - an air compressor assembly - and several modes of communication set up for experimentation. These are listed below:

1) copresence - both speech and vision, physically present,

2) telephone - verbal, but no vision,

3)&4) restricted communication - the set up described below was used to stop interruptions.

In addition, mutual visual access was restricted - only snapshots could be requested or given - and in 3) the apprentice was aware of the set-up, but in 4) the apprentice was told the consultant was really a computer. The diagram below in Figure 2-3 is taken from Deutsch (1974, p. 4) and also appears in Grosz (1977). The strangeness of Grosz & Sidner's dialogue example was traced to its origin in a corpus of specific experiments, 3) and 4) above, that were intended to eliminate interruptions in dialogue. In this particular fragment, the key point is that the dialogues were disturbed by a human mediator who interpreted the novice's spoken utterances and typed them to the expert on a Teletype; expert responses were typed and then read aloud by the mediator. In addition the participants could only see 'the same world' by explicitly requesting still snapshots of components; otherwise they were visually blind to each other's actions[8].

What is interesting about the example is where in this process did the dialogue fragment come from? Was it from a transcription of the spoken exchange between novice and mediator or an abstract record of the typed exchange between mediator and expert[9]? It is crucial to realise that even though it was a dialogue of sorts it cannot be claimed as an example that captures the essence of dialogue without the 'muddiness and unruliness' of talk on which to

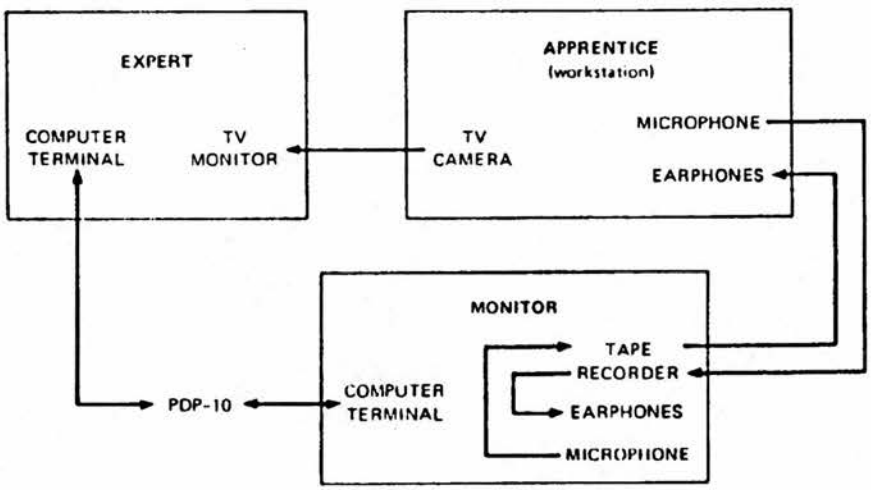
7 In the publication literature Deutsch became Grosz after this paper.

8 The keyboard and speech dialogues from these experiments were recorded and features of their structure and nature noted in Grosz (1977) and Deutsch (1974). Compared to Grosz & Sidner (1986) the earlier open pre-theoretical attitude is refreshing. Some of the phenomena noted are: interruptions, latched utterances, and conversational 'duets' commonly occur in experiments 1) and 2); and there is an attempted analysis of the marker "ok" in terms of sequential positioning; it is noted that it is possible to have multiple tasks - hypothetical and competitive - and misinterpretations.

9 The former is a possibility because in Grosz (1977, p. 14) it is claimed that "computer terminals were used solely so that transcripts could be easily obtained."

base a theory of human dialogue. Also, many phenomena have a new significance in this light, eg. the "OK" responses in lines 14 and 24 in Figure 2-2 may have occurred because of the explicit work required to coordinate access to a restricted visual display; the lack of interruptions is explicable not because of experimental shyness, but because it was not possible for the apprentice or expert to interrupt each other if the mediator did not allow it; the extended 'reflective' turn of lines 16-18 could have occurred because of the slow and erratic times of the speech simulator responses, that are out-louding the expert's typed response. In addition, many interesting and extremely relevant phenomena have been lost from the record, eg. the temporal delays and mediation strategies, and from the description of the experiment itself, eg. the mechanical techniques and procedures for constructing a speech or Teletype response. Later in Chapters 4 and 5 an empirical audio-visual study is

FIGURE 2-3 - A diagram from Deutsch (1974, p. 4) illustrating the restricted experimental setup.



reported that attempts to explicate important circumstantial features of a restricted modality dialogue. In addition in this case it would be very interesting to study the mediator's work of interpreting each participant's responses, and making the dialogue uniform and uninterrupted. How did the mediator manage the transcription of the apprentice's speech and the reading aloud of the expert's typed response? Did the mediator act as arbitrator on turn boundaries? That is, how did the mediator break up the stream of speech, and how and when was it sent to the expert? Did the expert decide on when his typed response would be spoken by conventional typing cues? Raudaskoski (1989) has shown, in a 'wizard of oz' simulation of a hypothetical telephone dialogue system, that the circumstances of interaction

are a crucial resource for the participants. For example, audio cues besides the speech stream and silence itself are significant for the achievement of intelligible dialogue.

Unfortunately, the data (tran-)scripts are already structured by interpretation of the conduct when rendering the analytic details, ie. the 'interchange' model is implicit in the practical reporting of communicative action. In this case, the transcript is either an automatic, temporally-abstracted trace of the mediator's translation to typed form and the expert's response, in which case they have a naive faith that the translation typed rendering achieved in the mediated interaction is somehow recoding those details that are relevant for the apprentice. Or else it is a verbal transcription of the speech exchange between the apprentice and the mediator. In either case it is not a record of what was significant for the participants in the interaction.

Grosz & Sidner suggest in their presentation of verbal examples that they be read much like one would read a dialogue in a story or a magazine, and this leads to an implicit claim that our intuitions and constructions of dialogue are adequate and faithful to real dialogue. So in fact it does not really matter what the script looks like as long as it looks like a 'dialogue' and makes sense - which is unfortunately misleading. The data clearly shows that the interactive circumstances are seen as a hindrance or problem in which the real core is the abstract interchange and recognition of intentions through language, ie. they separate structure and process from dynamic constitution and circumstance. Their analysis requires an ascription of intent and a segmentation based on the analyst's intuitions with no appeal to empirical justification outside of the manufactured verbal fragment. The validity of the model is in its prediction of those intuitive judgements without questioning the intuition themselves; what is required is an investigation of the participant's tacit knowledge implicated in dialogue practice.

Grosz & Sidner's empirical methodology is to present a verbal record of task-based dialogues and a number of intuitive examples. For example, there is a tendency to give data examples which are devised and imagined to be illustrations of everyday dialogue conduct, eg. (p. 192) "John came by and left the groceries//Stop that you kids//and I put them away after he left." From the problems described above, and given that the examples they use are restricted purposive dialogues with short turns and task-centred sequential organisation, it does not appear that a specific experimental task dialogue and a few intuitive examples are enough to base a theory on.

To conclude, the origins of Grosz & Sidner's data are not as they first appear. In trying to analyse their data from an interactional perspective it became apparent that the setting and interactive circumstances of the data examples were not seen as important or relevant to their theory. Their dialogues had been taken from a special experimental set-up that explicitly

avoided the possibility of interruptions. If the task and dialogue structure is the product of the orderly work of the participants in the local interactional setting, through practical resources, then to claim a general theory of discourse structure without any representation, notice or indication of these things is missing the phenomena it is claimed to deal with. It is not simply that their theory needs a better and more detailed theory of the process of coordination and recognition of intentions so as to fill out the abstract theory[10]. In the setting from which the example came, the circumstances of the interaction will be important at all levels of analysis and the precise details of their relevance cannot be taken-for-granted nor explained intuitively. What is required is an empirical study of dialogue activities in terms of their local and situated accomplishment and such a study is reported in Chapters 4 and 5.

2.6.4 Implementation and Artifact

What is the status of Grosz & Sidner's theory? It is not at all clear what could count as a successful theory, and what kind of data could be used as evidence for or falsification of the theory. It is hard to see how they could justify their intent or discourse structure ascription in terms of the data rather than an appeal to the intuitions of a fellow analyst. Also, it is not evident if the purpose of the theory is to give an intuitively satisfactory account of dialogue structure, or to provide the means to label and segment a fragment of discourse. In practical terms, it is puzzling just how a more concrete use of their model could be undertaken. There are two possibilities: each of the components could be taken as useful a 'knowledge source' that can be modelled in an implemented dialogue system so as to constrain interpretation and production; or an implementation that could automatically label and ascribe structure to dialogue examples that is faithful in some way to what can be found in human dialogue behaviour. But they have nothing to say about either of these alternatives. Also, the latter option does not validate the model as an accurate procedural model of dialogue as the structure may play no part in the achievement of the dialogue for the participants. It is unfortunate that the situated and circumstantial details of dialogue participation are ignored

10 For example, in Appelt (1985) an attempt is made to provide an account of the planning of actions by an individual required to support a conceptual theory of action for language generation. Situations are constructed to illustrate the cognitive and communicative processes of two face-to-face participants - an expert instructing an apprentice - who are attempting to collaboratively complete a task. The utterance that is used to illustrate the problems of planning a contribution is: "Use the wheelpuller to remove the flywheel". This scenario is remarkably similar to that in the data example in Grosz & Sidner (1986) derived from the original corpus used in Deutsch (1974). However, the participants are now face-to-face working over a table with a number of tools in front of them. It appears that the earlier example first entering the literature in 1974 has become decontextualised and recontextualised in a constructed example that perfectly matches the expectations of a reader of Grosz & Sidner's example in Grosz & Sidner (1986). This is an indication of the field's acceptance of examples into its 'repertoire for explanation' without regard for their circumstantial production.

in their theory for they are constitutive of the dialogue and not to be taken for granted or dismissed. These questions are hard to answer, but anyway it appears that Grosz & Sidner have developed a general theory that is quite comprehensive in its attempted unification of linguistic and non-linguistic aspects of dialogue, but they have not drawn upon the rigourously empirical work of conversation analysis. Chapter 3 describes the work of this field, and later in Chapter 6 its relevance to the conception, design and evaluation of computer dialogue systems is discussed.

2.6.5 A Brief Interactional Reconstruction

Briefly an alternative version of the basics of Grosz & Sidner's theory for dialogue activities will be presented. People engage in conjoint activities - lifting a large awkward table, walking down a street as a group, shaking hands, chatting - which can be glossed in an intentional language either before, as a projective plan for example, or after, as a retrospective rationalisation. Some of these activities can be seen, by the participants and analysts, to have certain relationships between them that are constructed in the discourse. In engaging in dialogue activity, participants demonstrate and accomplish this order in and through the dialogue. Accountable structures do emerge and are oriented to in dialogue activity. In particular, these structures are oriented to differentially. For example, an interruption is a recognisable achievement different from a side sequence. Accountable structures are accomplished as unique structures that are sequentially displaced and juxtaposed with each other. Certain sequential organisations can be displayed, or can be rationally reconstructed if necessary. For example, opening sections of telephone calls expectably come before closing sections - this is the sequencing or ordering relation. Also, side sequences or insertion sequences are expectably embedded in other activities and will return to them unless there is good reason - this is the constituency relation. Both relations have to be achieved in the discourse sequence, by the participants, contingently. However, they do not have to take the course expected, and they are interactionally reinterpretable, ie. interpretation is not once for all, as contributions can be located in different sequential environments through the course of a dialogue. Because dialogue and other activities are closely interleaved - participants are coordinating their bodies and talk in activities - it is understandable that the structure of dialogue is closely linked with the organisation of task-centred activity.

In the account above of the phenomena of interest to Grosz & Sidner and other models of dialogue there is no need to appeal to plans, beliefs, goals and intentions in order to explain the dialogue. Of course they can be used as part of a behavioural vocabulary - an intentional vocabulary is undeniably useful for accounting for our own and other's behaviours as

purposeful actions but like instructions they do not adequately specify action. Given this simple reconstruction, one path seems to be open and that is to take an empirical stance and study the orderly work that people do that constitutes the dialogue structure.

2.7 Summary

The problems and limitations of computer models of dialogue have been illustrated through a broad survey of the development of current models and a detailed examination of one particular comprehensive theory. Three crucial points brought out in this chapter are as follows.

First, computer models of dialogue have a weak theory of action and activity. Grosz & Sidner's theory does improve on previous computational accounts but is problematic for a number of reasons: the intuitive ascription of intent and structure; the plan-based theory of coherence; the lack of attention to processes for local constitution or improvisation; the avoidance of interactive circumstances and dynamics; and the proliferation of over-rationalised intuitive abstract representations. Modelling dialogue as the cognitive process of recognition and coordination of plans or intentions is unworkable in practice. Typically, computer models of dialogue are based on rationalised reconstructions and representations of action and structure which derive from observer's idealisations and intuitions that are used to generate and simulate dialogue behaviour. Interaction is not considered an important domain for constructing models.

Second, there are no strong phenomenological or empirical methods. An additional reason for taking an extended look at Grosz & Sidner's theory was to explicate how a weak empirical methodology has led to the elevation of a rationalised and sanitised view of dialogue. In their model the preconceptualisation of just what the phenomena is and what data is required is not satisfactory if details of the practice of engaging in dialogue are required.

Third, computer models of human dialogue provide a poor foundation for the conception and design of computer dialogue systems. There is an implicit separation into dialogue competence and performance, in which the interactive domain is peripheralised as something to handled later once a theory of dialogue and action has been understood, but, it is the claim of the thesis that the interactive and situated nature of dialogue is a crucial part of the development of an appropriate model. It is not that an abstract model will allow tailoring for each dialogue application. But it is the situated practices of achieving intelligibility in human activity, from which reconstructions and representations emerge, that are paramount in conceiving of appropriate dialogue artifacts. Four consequences can be found if these models are used: first, that structure and intention are ascribed without foundation; second,

that action itself is 'over-rationalised' so that the elaborated lay notion of planning replaces the circumstances of action; third, that those ascriptions are 'reified' as the basis for participant's work; and lastly, many emergent, tacit, mutual, local resources for action are unaccounted for and unavailable. These consequences are not trivial, they are fundamental misconceptions of the nature of dialogue and action that lead to impoverished notions of the nature and design of dialogue artifacts. An alternative vocabulary that is social in origin will be examined in the next chapter. Of course, aspects of computer models are extremely useful for the design of dialogue artifacts and interesting phenomena have been located and reasonably explained using the models. The argument is not against the usefulness of computational modelling in principle; for example, it may provide external 'etic' resources for dealing with dialogue conduct that are unavailable to people. Instead, in Chapter 7 it will be argued that an alternative use of computer modelling may provide interesting insights into the emergence of communicative activity.

An attempt is made in the thesis to bring in an alternative empirical and interactional perspective to inform future models. Chapter 3 will introduce the foundations of this alternative in hermeneutics, ethnomethodology and conversation analysis. The idea put forward in Chapters 4 and 5 is that studying human dialogue empirically will reveal features and phenomena that are constitutive of dialogue but have been missed by computer models. The experiment conducted was partly in response to Grosz & Sidner's dialogue example, and partly to study a modality that more closely resembles the possibilities available in design than the simulation of everyday conversation.

Chapter 3

THE PRACTICE OF DIALOGUE AND THE INTELLIGIBLE MACHINE: AN INTERACTIONAL PERSPECTIVE

3.1 Introduction

The last chapter considered contemporary computer theories and models of dialogue in terms of their ability to model aspects of human dialogue, their empirical validity, and suitability for the design of computer dialogue systems. In each case, the models were found to be problematic. In this chapter, an alternative interactional perspective on dialogue will be presented that supplies an understanding of dialogue as a collection of practices in which participants are mutually engaged in coordinating communicative action and achieving shared understanding. Rather than reconstruct interaction as the recognition and coordination of plans and intentions, the specific mutual achievement of interaction in situ, from which rational accounts emerge, is studied. This approach is derived from the empirically naturalistic fields of ethnomethodology, which investigates everyday natural activities, and conversation analysis (CA), which is particularly interested in the achievement of everyday conversation[1]. Their main concern is with the ongoing practice of comprehending events from the materials that make up the activity itself. The specific findings relevant here are: the documentary method, the indexical nature of language and description, and the methods of achieving intersubjectivity rather than the postulation of shared agreement.

The thesis will be concerned with the development of an understanding of practice and achievement of participatory dialogue itself that can inform the design of dialogue 'through' and 'with' machines. Additionally, given that computer models influence the design of

1 Both have a number of possible applications, eg. the achievement of work practice in and through artifacts in workplaces.

artifacts, the conception of dialogue and the machine's behaviour in the design of complex 'intelligent' systems will be examined. It is observed that the machine is intended for human consumption and to be intelligible, it is complex and reacts to human action with language forms. Thus, a shift has occurred in the conception of the machine from intelligent to intelligble. It is important to conceive of the dialogue machine as an interpreted artifact in human practice, and thus to understand the sense in which dialogue 'with' a machine in practice requires an implicit suspension of disbelief by a user in order to find the activity sensible. In addition, there is a clear sense in which designers of dialogue artifacts attempt to represent, and thus recover a semblance of, aspects of presence, referential practice and the living of language. The ambiguity between the machine as 'discourse medium' or 'visible author' cum social subject is also explained.

Using the methods of interaction analysis described here, a study of dialogue 'through' the machine will be presented in Chapters 4 and 5 that reveals some constitutive practices and features of CMCs as they are achieved as 'virtual dialogue spaces', ie. they are constituted as simulated shared graphic spaces in and through which dialogue can be conducted. This is undertaken partly to demonstrate what is ignored or unaccounted for in computer models of human dialogue, partly to study a restricted virtual dialogue space in its own right, and partly to suggest some resources and principles for the conception, design and evaluation of dialogue artifacts. The aim is to deflect attention from the over-rationalisation typical of computer models and suggest an alternative range of constraints and resources for useful computer models for dialogue artifacts. Later, in Chapter 6, interaction analysis is used again to study the integration of talk and body in activity, and suggest resources for coordinating and collaborating in dialogue activities interleaved with other activities and embedded in space and time. In Chapter 7, an alternative use of computer modelling based on attacking the fundamental problems discussed in Chapter 2 will be outlined.

The key issues for the remainder of this thesis, arising from the problematic computational modelling paradigm, are as follows:

i) A perspective on the dialogue machine.

ii) An interactional perspective on communicative action and practice.

iii) Empirical methods for focusing attention on the situated details of human or human-computer dialogue activity.

3.2 Situated Action and the Intelligible Machine: Review, Inspiration and Alternatives

The following sections are all concerned with understanding action, dialogue and the machine - how dialogue is conducted, achieved, and interpreted by people in everyday life and what part the machine can play in modelling or acting in practice. Major critiques of the traditional views adopted explicitly or implicitly by computational theories of behaviour and mind have been undertaken by Winograd & Flores (1986), Suchman (1987, 1988), and Agre & Chapman (1989), among others. All draw upon alternative perspectives from the human sciences of the nature of fundamental notions such as action, language, and interpretation. They also inform some of the arguments presented in the next sections.

We begin by looking at the conception of the machine in human practice, particularly in terms of inter-action and language-use. In order to develop an account, it is necessary to examine the practical nature and intelligibility of conduct that is to be undertaken by an implemented 'intelligent' machine. Rather than test or argue for the 'intelligence' of the machine, it is interesting to look instead at how the intelligibility of the machine relies on and is constructed by the user in situ. The discussion will begin from the early critiques of the 'intelligent' computer and move towards considering it as an interpreted and socially constituted artifact. The designer, the user and the reactive machine with its complex behaviour incorporating forms of language within the practices of its users need to be considered. Finally, a long introduction to social conceptions of practice and action that are driven by empirical concerns is given from which an interactional approach is derived and adopted in the remaining chapters.

3.2.1 Early Critiques: Intelligent to Intelligible

Understandings of intelligence, action, and language have all been influenced by the machine, and a number of disciplines or paradigms have arisen that use the computer to model and explain human behaviour. Critiques of the two main schools, artificial intelligence (AI) and cognitive science, have focused on the traditional rationalist assumption of rule-governed conduct, and the mentalist explanation of that conduct. Instead, the interpretative tradition in social science and philosophy provide an understanding of the machine and its limitations. Dreyfus (1967/72) was one of the first critics of the newly-formed field of artificial intelligence emerging from the fields of computer science and psychology in the 1960s. He has brought the names of Husserl and Heidegger, and the fields of phenomenology and hermeneutics to the attention of the cognitive sciences, even though the early debates were not very self-reflective. Dreyfus & Dreyfus (1986) continue the debate

with respect to the tacit and discursive nature of human practice that is claimed to be embodied as knowledge in expert systems.

Winograd's (1981) "What Does It Mean to Understand Language?" is his statement of a reconsideration of earlier work on natural language understanding systems of which he was an active proponent. Harris (1987) has also been critical of mechanical models of language that exclude the moral and normative aspects. Later, drawing inspiration from Dreyfus, Winograd & Flores (1986) and Winograd (1987) launched comprehensive critiques of the rationalist tradition that informs cognitive science and artificial intelligence. The work of hermeneutics, as well as the neurobiology of Maturana & Varela (1980), became the cornerstone of not only an attack but a way of thinking of the problem of designing machines as intelligible and not intelligent entities. In Winograd & Flores (1986) conclusions are drawn about the nature of computer science and design in relation to work practice. This has been explored further, for example, by Bodker (1989) in terms of the Russian school of activity theory.

3.2.2 The Interpreted Machine

3.2.2.1 Winograd and Hermeneutic Philosophy

Winograd & Flores (1986) have brought a distinctive vocabulary to the understanding of computers and their design and use, borrowed from the ideas of hermeneutics; particularly some ideas of Gadamer and also Heidegger in a brand of philosophy known as existential phenomenology. Hermeneutics arose from the exegesis of biblical texts and has developed into a collection of theories of interpretation and understanding. One fundamental claim of hermeneutics is that a text (or action) only has meaning in the activities of interpretation, and thus the objective reading of a text is impossible. Interpretation is a pervasive everyday practice in which an individual is continually involved in acting, understanding, and thinking.

Winograd & Flores read the difficult contributions of Gadamer and Heidegger as an alternative vocabulary[2] for explaining the condition of a computer user concernfully acting in the world and the problems engendered in designing artifacts for such a user. Some conclusions from their reflection on Heidegger in terms of the computer modelling of human conduct and understanding are as follows. First, we cannot necessarily make the 'tacit' or non-discursive explicit; thus representation is not a guaranteed enterprise resulting in exhaustive and explicit

2 The vocabulary is only derived and objectified for Winograd & Flores' purposes - it is not the translated vocabulary of Heidegger, eg. Heidegger (1967/27).

rules governing conduct. Second, practical understanding is primary and not to be subordinated to theoretical reflection. Action in the world is articulated through practical involvement in which artifacts are 'ready-to-hand' and not explicitly represented. Only in 'breakdowns' do artifacts or actions become 'present-at-hand' and represented. For example, the pen in my hand when engaged in writing is 'ready-to-hand', it is used unreflectively, but when the pen runs out it becomes 'present-at-hand', an object for repair, a 'pen' which does not work, or is playing up. The point is that representation only emerges from concernful practical activity, and the representation is not necessarily a part of that conduct in situ for the participant.

Unfortunately, Winograd & Flores somewhat mar their insights by resorting to a variant of speech act theory to explain 'language as action' and to build a system that supports computer-supported work practices involving communication. An alternative conception of action is presented in Section 3.2.3. Mallery et al (1987) have extended the debate by broadening to the work of other more contemporary authors, like Riceour and Habermas. The phenomenological hermeneutics of Riceour will be used in the next section to shed light on the nature of dialogue artifacts as part of the perspective on dialogue and the machine.

3.2.2.2 The Dialogue Artifact

> *In focused interaction participants engage in what Goffman has distinguished as 'interchanges'. In an 'interchange' it is common that first one person does something and then another does something, but these successive doings are treated by the participants as being somehow linked together... a characterisation of the principles that govern the way in which the succession of 'doings' in an interchange are linked together and how, as a result, they are organised into a coherent unit... attend to each other's behaviour in a highly differentiated way.* (Kendon 1988, p. 31)

What is the difference between a person talking with another person, a person using a complex computer system, a person walking a dog, and a person cooking? Consider this question in terms of the nature of the 'interaction' between the person and the objects implicitly engaged with in the activity. Does a person interact with a dog in the same way as with a person, or the egg being cooked, or the computer expert system? In some sense it can be said that a sort of 'interaction' occurs, an influence or reciprocal relation between the actions of one and the observable behaviour of the other, because the verb 'to interact' has many connotations and applications much like 'dialogue'. People's direct engagement and experience of the world is crucial for situated action and it is interesting how the actions of a person and the observable events in the world come to have a significance for that person. However, walking the dog and cooking the egg are usually not considered as social activities in the same way as dialogue or communicative interaction between people. On the other hand, technology has developed to the point where the significance of some machines'

behaviour is to be accounted for in terms of its intelligibility as dialogue or mutual communicative action. But it is problematic just how the actions of people and the behaviour of machine are linked as if in interaction, interchanging and attending to each other in successive doings of the machine and user and whether or not this can appropriately be called dialogue.

The 'dialogue partner' perspective[3] has been explicitly discussed in Kammersgaard (1985) in contrast to the tool, the system, and the media perspectives of the use of computers.

> *When applying this perspective humans and computers are regarded as partners in dialogue. The interaction process is regarded as a communication process in which user and computer application act as both sender and receiver, and the computer application is seen as being able to show communicative behaviour similar to that of its partner.* (p. 13)

Also, metaphors of person have been used to give perspectives on the different types of interfaces, eg. direct manipulation of icons such as in the Macintosh microcomputer are examples of first personness, and third person interfaces have processes mediating between the possible actions in the domain and the user that give the appearance of willing 'agents' carrying out your instructions if they are deemed safe. The names of books like "The Articulate Computer" by McTear (1987), "Getting Computers to Talk Like You and Me" by Reichman (1985), and the subtitle "Towards Conversational Computers" to the book by Waterworth & Talbot (1987), all lend to the idea of a computer and user in dialogue, particularly with dialogue in the ultimate form of conversation.

How did the 'machine as dialogue partner' metaphor arise? First, the beginnings of the metaphor are apparent in the emergence of the 'interactive computer'; instead of the computing machine that computes uninterruptedly given a program and a set of inputs, users could interrupt and change the flow of control in the machine at the time of execution. It was a simple step to move from the switches and dials of early models to easily reproducible and recognisable graphic character symbols, and then to stock textual language which became crucial as a means, for example, of informing the user of progress and asking narrow 'yes' or 'no' questions of the user that required a response at the run-time of the program. The sophisticated user interface had arrived.

Let us briefly consider how the metaphor developed in two closely related fields: the practical development of computer science, and the stormy progress of artificial intelligence. In

3 A reaction against the restricted metaphor can be found in Goranzon et al (1988). They reinvest the concept of dialogue with the richness that has been lost by its appropriation as computer jargon.

computer science, Licklider (1960) suggests that man and computer could work together, by moving away from the idea of a non-interruptable computation of a determined procedure to the notion of shared load, between machine and human, of problem solving. He claims that "human brains and computing machines will be coupled together very tightly and that the resulting partnership will think as no human brain has ever thought." (p. 4) The 'partner' metaphor seems to arise because of the ambiguity of determinate rule following in a procedure and the 'lived work'[4] of problem solving in program design. That is, procedures were to be executed as logical symbol manipulations embodied in the computer, and the process of refinement and correction of the algorithm by a human problem solver is iterative over successive full executions. However, the possibility arises, apparently, of the computer following the steps of an algorithm in the same way as a person might do, and therefore 'living' the work of the programmer or human problem solver. At appropriate points in the execution the computer becomes not a blind rule follower but a problem solver. Licklider foresees the problem solving activities "in which the computer cooperates, turning up flaws in the reasoning or revealing unexpected turns in the solution." (p. 5)

Martin (1973) uses the term 'man-computer dialogue' to refer to what happens when the user conducts some activity with or through the computer interface, screen, sound, etc. that has since become common currency. A later contribution by Nickerson (1977), that presages some of the work in the thesis, concerns the conversational metaphor for human-computer interaction. He gave a list of features of conversation that may be suitable and which includes: bidirectionality; mixed initiative; sense of presence; non-verbal communication; intolerance for silence. The list, including many other items, is now ripe for re-evaluation given the rapidly developing new technology and conversation research in the last ten years. For example, consider the item 'rules for transfer of control' which is representative of the signal view of turn-taking in speech exchange systems in contrast to the local opportunity management of CA. (See Wilson et al (1984) and Section 3.2.3.3.)

Second, the flowering of the metaphor can be found in artificial intelligence research from the 1950s that has been mentioned in the last chapter. Turing (1950) first raised in detail some of the problems for a computational theory of mind and how it could be judged to have succeeded. In an all too well known paper he suggests the possibility of an English language-using machine: "it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English." (p. 460) Also, the notion of a conversing machine is used in the familiar operational Turing test in which two people

4 See Livingston (1987) for an investigation of the 'lived work' of mathematical theorem proving.

and a machine play an imitation game such that an interrogator must identify correctly which of the other two is the human and which the machine by questioning through a reduced dialogue medium that restricts the mutual access of participants to that which cannot obviously determine the identity of the participants. Conversation is chosen because language use in talk is a highly skilled accomplishment. A successful conversing machine could be taken as a clear indication of intelligence. Unfortunately, the observation that people judge the intelligence of others from conversational activity is extended too far in the operational test; they do not routinely judge if someone is intelligent or not in conversation. The possibility of machines engaged in dialogue is also entertained by Katz (1966), in "The Philosophy of Language". The separation of understanding from situated conduct in which observable features of human conduct appear is demonstrated in the hypothesized simulation of

> *two giant computers with no ability to formulate thoughts and ideas or understand them but with the ability to produce speech sounds alternately in such a fashion that verbal exchanges between the computers replicate the publicly observable phenomena that occur when human speakers communicate in a natural language.* (p. 99)

It turns out that language use could be one of the least successful discriminatory tests, because people find sense in most cultural products in which they do not have access to the shared production of, and even when they do, the significance of actions can be indefinitely enumerated, as Garfinkel (1967/84) and Weizenbaum (1966) have shown.

The development of implementations that have tried to demonstrate a working dialogue machine, and thus support the 'dialogue partner' metaphor, has already been mentioned in Chapter 2, but here is a brief summary. Weizenbaum's ELIZA program is reputedly a conversing Rogerian therapist, that has been judged a skilful analyst by independent judges from transcripts of the dialogue between a patient and the machine (Weizenbaum 1966). Also, Terry Winograd's SHRDLU is famous as an early attempt at complete coverage of a micro-world consisting of toy blocks (Winograd 1972). More recent approaches are the expert systems of last 10 years that mainly use written forms of language, and the not very robust speech systems of the last 5 years. However, the development of 'intelligent' machines has resulted in a misuse of the vocabulary of mind and action. Many systems once called interactive data processing systems, now have the titles: expert system, knowledge engine, etc.. Equally, claims are commonly made about the relation of the programs and their representations to the mind. But the machine-centred view that regards the machine as an active partner able to deal with the contingencies of the human practice it supplants is now giving way to a more human-centred perspective. For example, parallel to the rise of the notion of the 'intelligent' computer and a computational theory of mind, Suchman (1987)

has coined the useful term 'interactive artifact' that focuses attention on the interpretation of machines in the activities of users, intelligibility instead of intelligence, and that the machine is designed for human consumption and embedded in human practice, much as a hammer or bicycle is. Machines can produce events in time, and 'sense' changes in the physical world, and thus, can be reactive to that world. They can be designed to produce fragments of behaviour, language forms, and they can be reactive to human use. The question arises: can the behaviour of a reactive artifact in the environment of human action be interpreted as meaningful with respect to the actions of people - ie. as social action? Suchman argues that because the computer artifact is reactive, language-rich and complex for the user, these are good reasons to consider it 'purposeful' in our everyday vocabulary.

Thus the properties of the machine in use give fuel to the argument that participants find the machine interactive and purposeful in their interpretations of its complex behaviour. Of course, this may not be appropriate for successful use but it does begin to explain the origin of the dialogue partner metaphor. People are very good at finding the sense or purposefulness of events in the world. Anyway, given that the machine exhibits language forms in its behaviour, it is interesting to raise the issue of whether or not it is language-using. This may be considered by comparing the similarities and differences between the 'language' machine and texts. Ricoeur has formulated a stance on text from phenomenological hermeneutics that is quite revealing in terms of human-machine interaction.

> *Only discourse not language is addressed to someone... But it is one thing for discourse to be addressed to an interlocutor equally present to the discourse situation, and another to be addressed... to whoever knows how to read.... The co-presence of subjects in dialogue ceases to be the model for every 'understanding'.... In escaping the momentary character of the event, the bounds lived by the author, and the narrowness of ostensive reference, discourse escapes the limits of being face to face. It no longer has a visible author. (Ricoeur 1981, pp. 202-203)*

Riceour follows in the tradition of hermeneutics in developing a sophisticated theory of the writer, text, reader and understanding. But the interactive computer presents a different form of text that poses a problem for the notion of 'author'. At present all systems are designed by people; they write the programs, and the language form, for instance written text or speech, is predetermined within robust limits for successful human use. Also, the language is discoursial as it is addressed to a 'someone', a user. However, unlike the presentation of language in most texts, the machine has to present appropriate pieces of language form in the context of the machine's use at the time of its use, ie. the momentary character of the event, and the referential circumstances of the interaction are now relevant. To put it another way: work on the interactive 'dialogue' machine is rediscovering the nature of copresence, of living language and of 'going on', in order to present appropriate linguistic responses to

the user's actions. The consequence of this is that the designer of a system cannot possibly know beforehand all the interpretations there will be as a result of the machine's situated use for the pieces of language she or he has decided to include there. Thus, the meaning of the language still escapes the bounds lived by the designer-as-author and so the ambiguity between the machine as 'visible author' or 'discourse medium' (authorless) is now clear, ie. the design of interactive dialogue artifacts is attempting to recover procedurally what it is to be a 'visible author' of language in practice when the language in discourse has broken free from the limits of occurring in human face-to-face interaction.

It is useful to compare this notion with Giddens (1987, pp. 100-1) on the cultural object or artifact. He argues that cultural objects, like texts and electronic media, have the following characteristics:

i) distanciation of 'producer' from 'consumer',
ii) consumer becomes more important than the producer in the interpretative process,
iii) a) durable medium of transmission across contexts,
b) a means of storage,
c) a means of retrieval.

The characteristics are similar for the interactive artifact. However, a stranger sort of distance occurs because the artifact, in a sense, manages meaning in the absence of the 'producer'. Meaning escapes the design horizons in processes of user-machine interaction as the user attempts to interpret the artifact's responses, and so it is difficult to talk about the designer or author, the user and the machine in appropriate ways. New hermeneutic accounts have to be debated because of the dilemma created by considering the machine as a social object or subject. For example, the interactive artifact is not like a static text; it is dynamic and reactive. But yet, it has qualities of texts in that the designer plays an important part as author, and there is a distance between designer and user. The interactive artifact must stand on its own, with no help from the absent designer, in moments of user engagement, ie. the artifact becomes a 'participant' in managing meaningful conduct. However, it must be remembered that the user sustains and grounds the appearance of an interface. An alternative, human-centred way of putting it is that the artifact must react in appropriate ways for the user, ie. the user finds it intelligible in terms of the project under way. In summary, this view of the computer explains the development of dialogue systems and models as the gradual progression towards recovering or replacing the referential qualities of talk. The dialogue artifact, as can be seen in the design of instructional computer systems that can be contrasted with textual materials, is reclaiming in small simple steps what the written language has lost, though it has gained in other ways, by suspending reference and divorcing itself from the

modalities of everyday experience[5]. Let us explore this example further as it helps reveal something of the shift involved from texts to computer systems. So what are the differences between instructional textual materials and reactive artifacts, like computerised 'intelligent' help systems[6]? Instructional texts are written for a number of reasons including the unavailability of a qualified instructor. Written and graphic documentation is produced that is to be read at some later date in order to understand, repair, build, or locate a device. It is advantageous in that it can be carefully designed and tested, and also re-read and mulled over, but is disadvantageous in that it cannot replace the embodied experience of apprenticeship with the qualified instructor (Brown et al 1988, and Lave 1988).

A simple observation is that though the language elements in both textual documentation and computer help systems could be identical, and the inadequacy and indefiniteness of both in specifying action is the same, a reactive artifact attempts to reclaim a part of the production of intelligibility. Texts are available to be read by choice with reader reflection on the consequences of a choice - an interpretation is constructed by the reader's work of looking and finding nexts to read or re-read. However, in reactive artifacts elements of the practice of finding a 'reading' are replaced or removed by the artifact. For example, rather than using meta-language or convention - a "Go to page 9 to read more on this" or a footnote marker - to make more explicit how to read the text, presentation of textual forms can occur without the user finding that next piece intelligible as the next thing to read. Finding a 'reading' from the flat textual page and the format of the book as a physical object are part of the routine practices of reading and interpreting, eg. searching for a page, scanning the text, re-reading, looking for a footnote. With 'intelligent' help systems much work of situated 'reading' required of the reader is to be replaced by the machine's operation: the model for interpretation is shifted from the monologic, that puts the emphasis on the consumer, to the dialogic. Meta-questions may explicitly ask the user for answers bearing on the presentation of material in a computer help system, but the user no longer has to engage in practices of reading an order into the textual materials such that a meta-question is found that suggests possible courses of action that must be carried out by the reader. In texts the guidance of 'reading' may involve meta-language, but the next courses are to be undertaken by the reader from the instructions in the textual medium, ie. courses of action must be juxtaposed in the text. However, in computer help systems presentation may be ordered in ways that are

5 In the sense of: "Meaning and reference are ordinarily closely combined in talk, not because talk is in any way primarily oriented towards description, but because it is carried on and organised within practical contexts of action." (Giddens 1987, p. 104).

6 Cawsey (1989) is an example of a prototype 'intelligent' help or tutoring system. Also, see Suchman (1987) for a study of the use of an expert help system for a physical device.

unobservable or unpredictable for the user, eg. the system models the user and presents according to these constructed models.

For example, if the explicit manual instruction "if it works then do..." were to be adopted in the design of a computer help system then the designer may have the machine determine satisfaction of the condition 'working' if the system detects certain states of a device and takes them as representative of the appropriate condition, and thus the contingencies of this rule will be hidden from the user. Also, let us consider the meta-question "Do you want to read more on this topic?" In a text manual the reader must find this as a relevant question within the context of the page and instructions as to how to proceed from the question to doing something sensible as a consequence must be provided nearby the question. The reader must then work out what about the text suggests a possible course of action and decide what course to take at this point, eg. "If yes then go to page 5", and then try to carry out the instruction. In a help system the question may appear at particular junctures in the use of the system in which it must be interpreted as a meta-question. What the user must do next is answer the question whereupon the system will hopefully provide an intelligible response, eg. if the user responds 'yes' then appropriate text (on page 5) will be provided. The possibilities for alternative action may not necessarily be presented to the user as they must be for the reader of a text.

Interpretation through 'reading' as a chronic and routine practice over time - a sort of interaction between the reader and the textual material - is transformed as the author or designer can remove some of the situated work of reading that is to the advantage of texts as distanced cultural products. That is, the design of interactive artifacts is recovering some of the situated practices of dialogue as the means to generate intelligible language material in situ. Now the sense of the language is to be found in the reasons for a particular next following a particular action, not in why a next follows a prior piece of text in the traditional linear and transitive order of a text. There is an endogenous display of interpretation in a next turn, that can confirm the appropriateness of prior action in instruction giving.

Because the design of interactive artifacts in the machine-centred view means recovering situated practice it is useful to think of the designer or author in terms of the following quotation from Perec (1987/70, p. xvii).

> *...despite appearances, puzzling is not a solitary game: every move the puzzler makes the puzzle-maker has made before; every piece the puzzler picks up, and picks up again, and studies and strokes, every combination he tries, and tries a second time, every blunder and every insight, each hope and each discouragement have all been designed, calculated and decided by the other.*

The designer must puzzle out the trajectories of possible user conduct and requirements, but in this case the puzzle is not static. System performance must provide a coherent and consistent picture of the activity engaged in for the user. This requires the intense 'lived-work' of programming and design. However, it is not possible to foresee every possibility, nor anticipate every user blunder. Trying to design the machine in this or any way leads to blindness and breakdown that is unavoidable in the circumstances of situated action. In this case the task becomes understanding the nature of design horizons:

> *the main tasks of the study of... cultural products of any kind... must be precisely to examine the divergencies which become instituted between the circumstances of their production, and the meanings sustained by their subsequent escape from the horizons of their creator or creators.* (Giddens 1979, p. 44)

Another interesting issue that relates to the question beginning this section is if the machine is interpreted then how do users of interactive artifacts make sense of it? There are several ways to address this question depending on your interests. For example, artifacts are encountered in specific locations and activities, they are encountered over periods of time by an individual, and they are encountered in the context of communities of practice. For the purposes of the thesis only the first aspect will be addressed. With what resources and interpretative strategies does a user get on with an interactive artifact? The designer hopes that in each reaction of the machine, the user will find the materials for interpreting the significance of that reaction. Let us consider a typical language-based help system. The computer reacts to the user with behaviours composed out of chunks from the top down, eg. the system can output complex linguistic segments in simple orders, like chunks of spoken text combined and altered in simple ways. The computer 'finds sense' in the primitive actions of the user from the bottom up, eg. the user can input simple segments in simple orders, like words in simple combinations, or apparently complex segments in simple orders, like a language-formulated stereotyped command. What counts is that the designer cannot work out the contingencies of complex combinations, but can rely on the user to ground a given pre-designed complex segment given by the system. Because of the linguistic and reactive nature of computer systems, through engineering and design, Suchman (1987) has argued that users find sense in the artifact using the resources of interaction to be found in human conversation. For example, what does it mean to take a linguistic form on a display as a question, and how is it different from a question in conversation? Sequential placement and the setting and activity are all relevant for interpretation. But, with an interactive artifact, the 'question' is a one dimensional exploitation of the regularities in language activity and the documentary method of interpretation. In addition, it is not at all clear what slippage there is between practical reasoning resources and social communicative resources, and how these

resources develop in the practical contexts of engaging with the artifact, ie. the similarities in the ways one interprets the physical and social world.

To conclude, in human-computer dialogue the computer can be roughly described as a 'dialogue partner' in the following terms: 1) it is reactive to human actions - when interpreted in terms of human action and activity; 2) language form is present, eg. writing or synthesized speech, that is to be interpreted in situ; 3) it is a complex system with no visible human author. This is of course the user's interpretive framework and not the machine's.

3.2.3 Situated Action and Empiricism

3.2.3.1 Introduction

This introduction to the study of social action and practice wants to show that there are traditions that are different to the rationalist conception of action, that can form the basis of an interactional perspective to current dialogue models and artifacts. Ethnomethodology provides the background for a reconception of dialogue that undermines computer models, and conversation analysis provides an empirical methodology that makes available the details of situated action that are otherwise taken-for-granted, ignored or lost. Dialogue can then be understood as a set of social practices loosely defined as the participants' communicative engagement in the interactional achievement of mutual intelligibility. The study of dialogue is concerned with how participants find sense in each other's actions with the circumstantial and situated resources of that activity.

3.2.3.2 Ethnomethodology

> *I use the term 'ethnomethodology' to refer to the investigation of the rational properties of indexical expressions and other practical actions as contingent ongoing accomplishments of organised artful practices of everyday life.* (Garfinkel 1967, p. 11)

> *The central recommendation is that the activities whereby members produce and manage setting of organised everyday affairs are identical with members' procedures for making those settings accountable... when I speak of accountable... I mean observable-and-reportable, ie. available to members as situated practices of looking-and-telling. I mean, too, that such practices are an endless, ongoing, contingent accomplishment.* (ibid., p. 1)

These two quotations from the founder of the group of studies known as ethnomethodology illustrate the difficulty a reader can have in understanding the vocabulary and the substance of its claims. The first step on the path to understanding is to recognise that the term 'ethnomethodology' is formed from the root 'ethnomethod' and the suffix 'ology' to denote the study of ethnomethods. Ethnomethodology originated in opposition to the traditional

sociological schools of thought in the mid-1950s and its stated aim was to investigate "the body of common-sense knowledge and the range of procedures and considerations by means of which the ordinary members of society make sense of, find their way about in, and act upon the circumstances in which they find themselves in." (Heritage & Atkinson 1984, p. 4) Garfinkel is the originator of the group of studies under this name and there are several classic analyses in his book Garfinkel (1967). Heritage (1984) contains a readable in-depth discussion of Garfinkel's contributions, and Sharrock & Anderson (1986) provide a similar broad brushstroke. Other useful sources are Lynch et al (1983), Benson & Hughes (1983), Heritage (1987), Livingston (1987), and Garfinkel (1986). Ethnomethodology[7] is representative of the confluence of approaches which Coulter (1989) calls 'epistemic sociology' concerned with the intelligibility of phenomena in practical affairs as a social achievement. Giddens has argued that it is also representative of a 'linguistic turn' in social theorising. However,

> *the 'linguistic turn' is in a sense a turn away from linguistics, conceived as an independently formed discipline, towards examining the mutual coordination of language and praxis.* (Giddens 1987, p. 80)

There has definitely been a turn away from cognitivism towards explanations that deal with human activities, actions and their circumstances rather than an internal mental world. Suchman (1987) has drawn on ethnomethodology in an important critique that argues against traditional views of action in cognitive science. Also, Agre (1988), Norman (1988) and Lave (1988) are some examples.

> *What we call cognition is in fact a complex social phenomenon.... 'Cognition' observed in everyday practice is distributed - stretched over, not divided among - mind, body, activity and culturally organised settings (which include other actors).... explanations of cognition as a nexus of relations between the mind at work and the world in which it works.* (Lave 1988, p. 1)

3.2.3.2.1 Theory

In sociology, Garfinkel moved against the traditional view that social order and coordination are the consequences of the application of sets of rules or norms that emanate from external social structure. Rather, social actors are "involved in a continued ongoing practical process of comprehending or making sense of the nature of scenes and activities... from the very

7 It is important to note that the field is fragmented in several ways, not least because of the peculiar nature of ethnomethodology, in that a fail-safe method or general approach is difficult to define because methods emerge from the specifics of each activity. Also, there are many approaches that can come under the rubric of 'investigations of everyday social life', eg. Coulter (1979), Goffman (1981), Cicourel (1973), and Douglas (1971), though they could not be said to present a unified perspective.

materials that they see as making up these environments." (Lee 1987, p. 23) Also, the roots of ethnomethodology go back to Husserl and phenomenology. In common with phenomenology's interest in everyday perception, the phenomena to be studied were not hard to find (unlike searching for black holes in space.) The naturally organised ordinary activities of everyday people were of great interest but had unfortunately been taken-for-granted in mainstream sociology. The order to be found in these activities was now made problematic and to be explained, but not as the result of the instantiation of social norms, nor as the result of common cognitive structures.

In this brief survey three basic points will be discussed as they bear on the problems with computer models of dialogue discussed in Chapter 2: how participants find sense in the world and other's actions using the documentary method; the indexical nature of language; and the mutual intelligibility of communication as a situated achievement. The 'documentary method' or 'verstehen' is a principle process of sense-making - it describes the process whereby action appearances are treated as 'the document of' an underlying pattern, and equally, the individual documentary evidences are interpreted in terms of those underlying patterns, in mutual elaboration. As Suchman (1987, p. 64) illustrates: "the ascription of intent on the basis of evidence, and the interpretation of evidence on the basis of ascribed intent" are abilities that are not reducible to formal relations between context and intent.

> *Interpretative schemes form the core of the mutual knowledge whereby an accountable universe of meaning is sustained through and in processes of interaction.* (Giddens 1979, p. 83)

The routine and pervasive nature of this method was demonstrated by Garfinkel in an experiment involving advice counselling between an experimenter and students in the University. They were only able to communicate by intercom, and the students were told that the advisor would answer their questions either "yes" or "no". Students were asked to comment after every response with the intercom switched off. In fact, the experimenter answered either way according to a random procedure. However, this did not perturb the students who found sense in the responses as motivated by the questions and addressing the substantive content of those questions. The documentary method thus explains why artifacts can be used. Fundamentally, it is people who make sense of the behaviours of animals, cars, and computers, even when those behaviours are unmotivated by their actions.

Garfinkel was very much concerned with natural language use and its relation to social practices. A fundamental claim is that all language is indexical, and that participants are actively engaged in making definite sense of indefinite descriptions or indexical expressions by invoking resources such as context or commonsense knowledge. It is true that

> *ordinary language is 'open'... most of the words and phrases used in everyday talk do not have precise lexical definitions.* (Giddens 1987, p. 102)

But, settings of talk are used by participants to define the nature of what is said. Heritage (1984, p. 151-2) illustrates the indexical nature of language with the example utterance "That's a nice one" said in front of a picture. He concludes that in the example: i) the referent could not be determined without physical (verbal) context, eg. 'that' in the picture; ii) the sense of particular expressions could not be made without the use of context, eg. 'nice'; and iii) the sense of the utterance (construed as action) could not be made without invoking a social context coordinated with the sense of particular descriptive terms, eg. a compliment. Thus, in this way, plans, such as in computer models, are indefinite representations abstracting away from action and are indeterminate of action because they are necessarily incomplete. The determination of the objective conditions and consequences of action or intent recognition is an endless task not belonging to the world of action itself. Also, rules do not govern conduct; they can be used to make practices accountable but not explicit. The application of rules involves the use of ad hoc devices, like et cetera clauses.

> *'Shared agreement' refers to various social methods for accomplishing the member's recognition that something was said-according-to-a-rule and not the demonstrable mapping of substantive matters. The appropriate image of a common understanding is therefore an operation rather than a common intersection of overlapping sets.* (Garfinkel 1967, p. 30)

It is crucial to understand that the intelligibility of action is constructed in situ according to tacit methods - the documentary method - of sense making. To reiterate an argument in Chapter 2, it does not rest on a shared agreement about background knowledge.

> *Situational constitution is essentially a 'local' and immanent product of methodic procedure rather than a result of 'pre-existing' agreement on 'matters of fact'.* (Heritage 1984, p. 132)

Therefore if it is methods and not mappings that underpin the achievement of mutual intelligibility then it is clear why plans and representations fail to deal with situated action.

The normative order was demonstrated by Garfinkel (1967) in his famous 'breaching' experiments in which students were asked to engage in conversation with a friend or acquaintance and sublimely ask at some point for the other person to clarify the sense of a commonplace remark. For example, if a friend said, "I had a flat tire yesterday", the student replied, "What do you mean, a flat tire?" The students reported stunned, perplexed and even hostile reactions in which the friend attempted to return to the 'normal' conduct that had been upset or breached. Thus the experiment demonstrates the production of conduct is founded in the perceivably normal that is continually ratified and achieved. In addition, this experiment opens the way to the study of the organisation of action itself. If actions can be analysed in terms of constitutive structures and they are displayed in the organisation of action itself then the ordinary competences and structures are available to participants and observers by virtue

of being a member of that community. This is the prime resource for the analyst in revealing the competences implicated in members' practices, and is used in Chapters 5 and 6 with respect to audio-visual study of dialogue practice.

It is important to note that two strains of ethnomethodology are present in Garfinkel and continue on today. Cohen (1987) argues that activities do exhibit a formal structure independent of any given cohort of actors, but are produced and recognised as the practical, situated accomplishment of the members of a particular cohort. There are then two possibilities: either thematise the practices and procedures by which standardised features of forms of activity are produced which is the course of conversation analysis, or attention is paid to the production and recognition of these forms in particular situations which has been continued by Garfinkel's students. Conversation analysis will be described in detail later but is basically concerned with "the possibility of achieving a naturalistic observational discipline that could deal with the details of social action(s) rigourously, empirically, and formally." (Schegloff & Sacks 1973) The latter studies will be briefly summarised next after the following quotation which sums up the perspective towards situated action and structure taken in this thesis.

> *The social world is constituted by situated actions produced in concrete situations, that are available to the participants for their own recognition, description, and use as warranted grounds for further inference and action on those same occasions a well as subsequent ones. Situated actions are produced through context-free, context-sensitive mechanisms [for example, turn-taking] of social interaction, and social structure is used by members of society to render their actions in particular situations intelligible and coherent. In this process, social structure is reproduced as an objective reality that partially constrains action. It is through this reflexive relation between social structure and situated action that the transparency of displays [the mutual intelligibility of conduct] is accomplished by exploiting the context-dependence of meaning.* (Wilson 1983, p. 20)

3.2.3.2.2 Doing Ethnomethodology

In order to illustrate a number of points about ethnomethodology a summary of a typical investigation of human conduct will be given. Examples of everyday social activities are the encounters of strangers in public places, eg. standing in a queue, spitting in the gutter, asking the time, street walking, crossing streets at appropriate places. The latter is a skilled accomplishment carried out by a large cohort of people without much thought, though it is not considered a great achievement of humankind. The following account of its structure and achievement is taken from Livingston (1987) in order to illustrate a number of points relevant to understanding and investigating the nature of dialogue activity.

How should the study of the behaviour of people crossing streets be undertaken? What is to be explained? People stand on opposite sides of a traffic-ridden street at conventional

places and then at the appropriate signal cars stop and people move from their side in order to reach the other side safely. Preliminary observation shows that no simple rule operates to order the flow of people. Crossing is accomplished in varied ways and how it is achieved is in need of explanation. One approach would be to video from above the patterns of flow that emerge on real occasions of street crossing. By using video analysis and photos it could be argued that people form geometrical shapes - wedges, fronts - with their bodies and thus by their collective behaviour achieve a crossing. However one thing clearly remains to be explicated. How do the pedestrians manage the crossing themselves. It could then be hypothesized that people orient to these structures in order to achieve the crossing, otherwise how could they manage it. It could also be argued that people have a shared plan to form these structures by which they can coordinate their crossing. Or that they have been socialised to follow these rule-governed patterns. They may even have been born with these patterns in their genes.

However, an investigation of the achievement of the crossing needs to consider more closely the routine social accomplishment of this activity by the members themselves. How is it that individuals can cross a street in a multitude of other people and retain their anonymity? A detailed ethnomethodological account would have to investigate from the perspective of the participants in the cohort, for example, in terms of directed eye movements and body orientations. They are bodily and perceptually involved in an emerging social activity in which their practical action and reasoning is constitutive. Visually the action so to speak is not from a disengaged observer's position above, from where representations become derivable from the video records, but from the eye level of those engaged. Simply, during the ongoing crossing pedestrians are engaged in the social production and maintenance of an 'interface' between the two opposite moving groups of pedestrians. Livingston argues that "they are engaged in locally building, together, the developing organisation of their mutual passage." (p. 22)

This example nicely illustrates some important points. First, the accomplishment of everyday activities is studiable. Second, activities are the result of embodied and situated practices. Third, the access to the competences is only available to members themselves. The analyst cannot escape having to be a member in order to study a strip of conduct. The analyst and the participant are both relying on everyday interpretative methods in order to make sense of conduct. ie. "the description of human activities demands a familiarity with the forms of life expressed in those activities." (Giddens 1984, p. 3) It would not be so easy to reveal the sorts of competences implicit in the conduct if one was blind or paralysed, much as listening to a conversation in an unfamiliar language is insufficient for doing analysis. Fourth, video is not necessarily the provider of the correct data. It is problematic for locating what is

demonstrably real and 'visible' for the participants. This problem for analysis applies to all human activities and is to be determined for each individually. Lynch (1985, p. 7) argues that "in many settings it is the embodiment of speech and gesture which provides work with its visibility for practitioners." Morrison (1981) studied the sequential organisation of texts and devised formatted questions and states that "a strictly 'literary' analysis of such texts (poems, sociology textbooks) cannot assure the analyst of a grasp of the organisational use of texts within the specific occupational activities which produced them." Fifth, inadequate consideration of the data results in an abstract and disengaged "technical" analysis that can lead to reification of those results as elements of a generatory account of the conduct. Structures described from disengaged perspectives are posited as endogenous to the conduct. This has been argued in Chapter 2 with regard to computer models of dialogue like Grosz & Sidner (1986).

Other areas of study, from the enumerable practical activities that are available, are: scientific laboratory work, martial arts, the reading of signs, the production of direction maps, jazz piano playing, mathematical proofs, and typing. However, the next section introduces the ethnomethodological investigation of a pervasive activity that is now considered an area of in-depth study in its own right: everyday, mundane conversation. It is important because of

> *the ability to observe the process of social action, whereby the social world is essentially built up anew for the purpose at hand, and interactants can be seen sorting out the agreed-on premises from those that need to be established between them.* (Wynn 1980, p. 88)

Given the emphasis on the everyday competences which members rely on when they engage in social practices and the central place of language in accounts, it is understandable that routine conversation itself, in and through which much social behaviour is constituted, became the focus of great interest in the 1960s. However, at the time it was something of a surprise.

3.2.3.3 Conversation Analysis (CA)

Conversation[8] is a routine and complex accomplishment carried through by almost all members of society with great skill and transparent ease. It is not only carried out in and through speech but also in the sign language of the deaf community. However, the

8 There is a contrast between the everyday general use of 'conversation', ie. 'conversational activity' versus the speech event 'conversation' which can be characterised as talk-in-interaction maintaining equal speaker rights (Wilson 1989). Also, the term 'conversation analysis' is sometimes ambiguous between CA and the studies of conversation in general, including other approaches. Here it will refer to the specific field of studies.

investigation of this pervasive activity was largely absent from academic disciplines until the second half of this century. Before, only 'interesting' speech events, for example, of a rhetorical nature with important social or historical consequences, were dealt with, eg. debates or speeches which had associations with formal or written language. The term 'dialogue' has a higher status in this respect because of its associations with literature and meaningful, significant language use. The discipline one would have expected to have dealt with language use in its most natural form, namely linguistics, initially avoided the issue, claiming disorder and unruliness.

Instead, the complex social order of conversation was revealed in a very different fashion by an empirical discipline known as conversation analysis that grew out of ethnomethodology. At first, no particular priority was assigned to conversation as its subject matter. But, as the field progressed, mundane conversation was studied in its own right and in priority to other forms of language use. The central goal of conversation analytic research is the description and explication of the competences that ordinary speakers use and rely on in participating in senseful conversation. Meaning is not built into the codes of language as in structural accounts, nor is it to be found in the relation of cognitive representations to that which is represented. Meaning is constructed in situ by interactants who are

> *simultaneously engaged in fine-grained real time co-ordination of speaking turns tracked predominantly in terms of surface structural features and... organising their actions in terms of publicly accountable normative expectations bearing on the nature and design of their turns at talk.* (Heritage 1989, p. 26)

Of course, personal motivations, personalities, cultural conventions, etc. are not to be discarded. However, conversation analysis, ethnomethodology and other 'interpretative' approaches have had little effect on computational theories of human intelligence and behaviour, though this appears to be changing gradually with the wide dissemination of directly relevant work by Dreyfus (1967/72), Winograd & Flores (1986) and Suchman (1987).

In this survey of CA, the fundamental principles in relation to computer models of dialogue will be described first. The empirical methodology will only be briefly outlined as it is dealt with more thoroughly and in the context of the interactional analyses carried out in the thesis in Chapters 4, 5 and 6. Some findings and other similar social approaches relevant to considering dialogue activity from an interactional perspective will also be reviewed.

3.2.3.3.1 Principles

The work of Sacks, Schegloff & Jefferson in the 1960s has been taken up increasingly in Europe, especially in Britain. The original pioneering work of Sacks has been generally unavailable except for a few monographs transcribed by Gail Jefferson, and the original published papers. However, recently some of the transcribed lectures have been published,

Sacks (1989), and all of them will be available later in full. Useful summaries of CA can be found in Levinson (1983), Atkinson & Heritage (1984), Heritage (1984, ch. 8), and Button & Lee (1987). Also, Roger & Bull (1989) contains a comparison between the perspectives of CA and social psychology. The field has moved from an early interest in social category membership devices and the accountability of conversational activity to the formal description of regularities found in conversational data, in which details of the sequential organisation of talk became prominent. After twenty years of research and critical attention the field has achieved credibility and findings are being incorporated into other disciplines, such as social psychology, linguistics, and discourse analysis. The main tenets of CA are summarised in Heritage (1989, p. 22) as follows:

> i) interaction is structurally organised,
>
> ii) contributions to interaction are contextually oriented,
>
> iii) these properties inhere in details of interaction so that no order of detail can be dismissed, a priori, as disorderly, accidental or irrelevant.

The first tenet can be traced back to the divergence in interests between Garfinkel and Sacks about what ethnomethodology should explain. Sacks (1984, p. 21) states that the central findings of ethnomethodology are:

> *The detailed ways in which actual, naturally occurring social activities are subject to formal description. Social activities - actual, singular sequences of them - are methodical occurrences. That is, their description consists of the description of formal procedures persons employ.*
> *The methods persons employ to produce their activities permit formal description of singular occurrences that are generalisable in intuitively nonapparent ways and are highly reproducibly stable.*

Conversation analysis has since become highly technical and detailed, painstakingly revealing structures in talk that are accountably used by members of a society to construct and interpret that talk.

> *The initial most fundamental assumption of CA is that all aspects of social action and interaction can be examined in terms of conventionalised or institutional structural organisations which analysably inform their production. These organisations are to be treated as structures in their own right which, like other social institutions and conventions, stand independently of the psychological or other characteristics of particular participants.* (Heritage 1989, p. 22)

The second tenet states that communicative action is both context-shaped and context-renewing. Actions in talk cannot be understood independently from the contexts in which they were produced because

> *in ordinary talk, individuals routinely employ a diversity of aspects of setting in order to understand others and to 'gear' what they themselves say to such a process of understanding.* (Giddens 1987, p. 100)

In interaction, speakers and listeners rely upon a "saturated physical and social context for making sense of what is said." (Giddens 1987, p. 127) Also, communicative actions themselves provide more material for next interpretations and thus engage in and renew those same contextual frameworks.

The third tenet underlies an empirical methodology that must be susceptible to not-yet-known orders of structure. These orders are only revealable in the observable details of conduct itself by careful analyses of recordings. This does not mean that as much as possible must be observed and will be relevant empirically, as Oldman & Drucker (1985) misconstrue the assumption. Rather, details must not be dismissed theoretically as irrelevant, for example by coding the data, lest one get rid of the forest in order to see the trees more clearly. This is a common mistake in computer theories of dialogue as was illustrated with respect to Grosz & Sidner (1986) in Chapter 2. The naturalistic methods of CA will be returned to in Section 3.2.3.3.3.

3.2.3.3.2 Phenomena

Now we will turn to the consideration of some basic findings and their relevance to the thesis. In detailing interactional and communicative resources, note that a mentalist vocabulary is explicitly avoided. Two fundamental organisations of conversation are turn taking and adjacency organisation, and they will be considered in detail.

In order to counteract the tendency for the 'interchange' model of message exchange to be adopted in models of dialogue it is pertinent to consider just how the production of utterances is achieved in talk. The classic work, much referenced and misinterpreted, is Sacks et al (1978), which was first published in 1974. Turn-taking is a pervasive phenomena in any speech exchange system, and can also be found in any interactional activity in which turns at participating must be coordinated. CA has found a number of properties of conversation in this respect which are in need of explanation. For example, how is it possible that parties in a conversation can regularly achieve split second transfers between the termination of speech of one party and the start of speech of another? Parties engage in talk in which the number of parties may alter, the length of talk by one person is not predefined, and the order of the 'turns' at talk is not determined in advance. CA has investigated this phenomenon and found regularities in the way people achieve the orderly distribution of turns-at-talk. Before explicating more about turn-taking it should be noted that conversational participation is organised within the constraints of mutuality. In talk settings, the sense of talking together - of mutual access to social and physical circumstances - is a coordinated, routine achievement. It requires "the coordinated monitoring which copresent individuals carry on as part of ongoing talk." (Giddens 1987, p. 100) In addition, participants use whatever

resources are available in the circumstances, and thus, inherent in the participation in conversation and other activities are emergent events that are available as a resource for participants to accomplish the conversation (Wilson et al 1984).

To return to turn-taking in conversation, turns are produced, competed over and oriented to by the participants. CA proposes that people construct turns out of turn-constructional units (TCU) that project possible points in the course of the conduct of the speaker at which transition may occur. At the transition relevance places (TRP) a simple set of ordered options are recursively operative at each and all TRPs.

a) if a current speaker C selects a next speaker N, C must stop speaking and N must start at the TRP,
b) if C does not select, then another person may self-select,
c) if neither b) nor c) then the current speaker may continue.

These 'rules' are locally managed and are not generative of the turn-taking conduct. Turn-taking operates on a turn-by-turn basis organising the transition to next speaker with the materials at hand. The rules do not have to be obeyed but most conduct can be shown to be produced and interpreted in the context of such a set. For example, the system provides a motivation for organising conversation so that only one speaker speaks at a time, however, overlaps do occur and are not breakdowns nor unexplainable because they do not satisfy the set of rules. Overlap is an ordered phenomenon that is explicable in terms of orientation to the rule set. An overlap can occur as a competing first start, as allowed by rule 1(b), or as a mis-projection of a possible TRP, as in the example below.

[Sacks et al (1978, p. 17)]
A: Uh you been down here before // havenche.
B: Yeah.

In this example, material has been appended by A to expand a turn. Tags or address terms regularly accomplish this expansion and thus redefine locally the boundaries of the turn in progress. However, B anticipates the possible TRP after "before" and fits a reply into this slot. Overlap occurs, which is usually brief, and talk continues. Thus in the course of the construction of the talk competing analyses by the two parties are an inherent part of that talk. An analysis of overlaps will not reveal an underlying order or coherence that was intended but realised incorrectly, because order is itself a contingent circumstantial accomplishment. The turn-taking rule set does not govern the behaviour of people in conversation but is merely a description of the orientations to what is a 'turn' that must be routinely and tacitly achieved on every occasion. For example, collaborative turn completions, in which one utterance is completed by another party before a recognised

completion point, illustrate that participants must construct 'turns' interactionally and this can be shared by two parties, ie. conversation is fundamentally a joint activity.

The turn-taking rule set also can explain the significance of an absence of activity as well as the presence. Thus in the following example the two pauses are differentially interpretable within the developing talk.

[Sacks et al (1978, p. 25)]
C: Well no I'll drive (I don't mi//nd)
J: hhh
(1.0)
J: I meant to offer.
(16.0)
J: Those shoes look nice ...

The one second pause is a gap before the application of rule b) above. The second pause can be interpreted as a lapse on the non-application of the rule set. Only incipient talk, common at work or home, escapes the accountability of absences of talk activity, ie. two people working in the same office may fall into silent activity for long periods without either feeling uncomfortable.

Within the turn-taking of talk some relevance of turns to each other is necessary. Adjacency is a prime locus of intersubjectivity. Adjacent positioning provides a resource for determining the adequate interpretation of a speaker's turn by a recipient that would not be available if turns addressing other turns were regularly positioned many turns away. Non-adjacent positioning does occur in talk, but these occurrences 'prove' the systematicity of adjacency because such occurrences are demonstrably accounted for in terms of their non-adjacency. Thus, a speaker might say, "I'm sorry for not answering earlier but..." or "I won't reply until later." Adjacency pairs are one routine structure available for the production and interpretation of talk. They are an additional local resource for organising transition to next speaker within the turn-taking system. Production of first part of a pair - a question, say - sets up the expectation of a second part - an answer - and thus influences the interpretation of what comes next. Of course second parts may not be forthcoming or may be positioned much later in the talk, but that is not a disproof. The expectation of a second part is oriented to in the production of whatever conduct occurs and a noticeable absence can be sanctioned by the speaker if necessary. Thus, answers do not have to follow questions immediately, if at all, and thus they are conditionally relevant. Also, repair is a pervasive practice: talk has intrinsic mechanisms for the resolution of turn-taking troubles, eg. overlap resolution.

To conclude this section, it must be emphasized that the sequential and interactional production of talk is not governed by rules. Some models propose that institutional rules

govern the behaviour of those involved such that turns at talk are signalled, accepted and then transferred (Duncan 1974). Computer models are similarly problematic in that they aim not only to predetermine what is significant in dialogue but also that what is significant must be intended. Rather, the set of rules can be said to tacitly inform the production and interpretation of turns by parties engaged in talk. Kendon, writing about Goffman and the 'given' and 'given-off' in talk, notes:

> *It is important to remember that the issue here is not whether the information provided is in fact provided voluntarily or involuntarily. It is, rather, whether the co-participants in the gathering take it that it is provided voluntarily or not. In any situation of interaction, it seems, participants treat only some aspects of each other's behaviour as if it were deliberately intended to convey something. In conversations it is usually called 'content of talk' that is created in this way, not the manner of talk, and certainly not the bodily stagings and ecological arrangements within which talk is carried on. However, it is not as if these other aspects of the situation play no role in the structuring of the interaction. Far from it. Their role is crucial to the whole way the event is organised.* (Kendon 1988, p. 23)

Thus, work is carried out in conversation that may not be explicitly identified as intentional but still is crucial to the achievement of talk. The recognition of intent as a prerequisite of understanding is not required either. But participants may treat talk as intended, and this may involve reinterpretations. Giddens argues that:

> *Human action occurs as a duree, a continuous flow of conduct, as does cognition. Purposive action is not composed of an aggregate or series of separate intentions, reasons and motives.... Terms such as 'purpose' or 'intention' ....extricate human action from the contextuality of time-space.... 'Action' is not a combination of 'acts': 'acts' are constituted only by a discursive moment of attention to the duree of lived experience. Nor can 'action' be discussed in separation from the body, its mediations with the surrounding world and the coherence of the acting self. (Giddens 1984, p. 3)*

Thus discursive reasons are different from the routine rationalisation of conduct, ie. "the reasons actors offer discursively for what they do may diverge from the rationalisation of action as actually involved in the stream of conduct of those actors." (Giddens 1984, p. 4)

3.2.3.3.3 Empirical Methodology

Because of CA's third tenet discussed above, the analysis of conversations is strongly 'data-driven'. CA painstakingly studies how talk is methodically produced by members in orderly ways that exhibit that orderliness for one another. The methodology avoids the use of interviewing techniques, the use of field notes or pre-coded schedules, the use of native intuitions to invent examples, and the use of experimental manipulation or directing behaviour which may restrict the range and authenticity of the activities which are elicited. Instead, the interest is in revealing the

> *organised procedures of talk as they are employed in real worldly contexts between persons in real relationships whose talk has a real consequentiality and accountability.* (Heritage 1989)

Some record of the phenomena is essential in order to study it in detail. However, the record must be of the details that were meaningful for the participants, and constitutive of the activities under investigation. CA has shown that audio and video tapings are adequate for investigating conversation on the telephone and in copresence. A recording allows repeated viewing of the event and consequently peer group ratification of a finding. Also, an impressionistic transcription system has evolved, that is now quite complex (Jefferson 1989). For example, details of overlapping speech, particular speech production characteristics like pace and mid-phrase cutoffs, and split-second timings of pauses are notated. This renders the talk 'strange' and thus can expose the everyday accomplishment that is taken-for-granted.

Analysis can then take two directions: a qualitative analysis of singular fragments in great detail, or a distributional analysis of occurrences across many fragments of talk. A primary resource for the analyst is that in next turns participants display public interpretations of the current action for each other in the mutual construction of the intelligibility of the talk. Thus a competent observer or analyst has 'visible' evidence for the in situ work by the participants themselves constituting the sense of the talk. Instead of ascribing intent or structure on the basis of 'what you would expect' or 'that is what I would intend', analyses can draw on the treatments given in the talk itself. To reiterate a point made earlier, an intentional language of description and an investigation of speaker purpose is avoided in the work of CA.

> *And, although a shorthand 'intentional' language is employed in these chapters, the perspective focuses on the underlying structures informing the interpretation and treatment of a speaker's action by a recipient and maintains, except in specific cases, a relatively agnostic stance on the question of how far the speaker consciously aimed at some particular interpretation. This agnosticism is consistent with the proposal that the objects of study are institutionalised structures of talk that are oriented to by speakers with varying degrees of reflexive awareness.* (Heritage & Atkinson 1984, p. 7)

The early interest of CA was in everyday mundane conversation, particularly on the telephone, with a parallel interest in institutional talk in comparison to the general core findings about multi-party conversations. Later, in the last ten years, research has gradually located talk in the context of the body and other activities participants might be engaged in. This research is still in its infancy but suggests that language, body and perception are intricately interwoven in the interactive achievement and coordination of those practical activities (Goodwin & Goodwin 1987).

3.2.3.3.4 Mistaken Views from Other Perspectives

A range of criticisms have been made of CA, and especially of ethnomethodology. I will examine a few of them to illustrate further some of the principles. It is often argued that CA has some interesting findings but that they are not theoretically interesting and are uninformative. For example, Power & Dal Martello (1986) and Brown & Yule (1983) acknowledge the results - turn-taking and adjacency pair - but claim they are simple and theoretically they contribute nothing to a understanding for the analyst of how conversational function is systematically realised in linguistic form, eg. when an interrogative form might serve as a question. This demonstrates a misunderstanding of CA. First, the findings of CA may seem innocuous and uninteresting for theories concerned with power or ideology in discourse, but the local organisation of action is not irrelevant for the constitution of social structures. Second, theorising was initially abandoned by CA because of the proliferation of ad hoc categories and unjustified speculations. By taking a naturalistic approach, regularities and patterns that are not intuitive nor immediately apparent were revealed and built on. Third, CA cannot say anything about a general set of rules for mapping form onto function because it is a local and circumstantial matter in which conversational sequence is a crucial resource (Sharrock & Anderson 1987)). Linguistic form is not to be neglected or forgotten in analysis, as Schegloff (1979) demonstrates, but it cannot be addressed in terms of the traditional 'linguistic' methodology.

There are also problems in understanding fully the notion of 'adjacency'. It is not a claim that all talk is to be interpreted only with respect to the adjacent placement of responses, or only as the exchange of adjacency pairs. Schegloff (1988) argues that Goffman made this latter mistake. Adjacency is a resource for finding the sense of contributions in conversational interaction which does not have to be obeyed nor does it always have to be satisfied by producing an adjacency pair. Another criticism from those concerned with computational and procedural models is that CA does not provide an adequate detailed account of how "coherence and sequential organisation in discourse is produced and understood" (Levinson 1983, p. 286). For example, language and conversation structure has become topical recently. This movement has unfortunately been taken up by some disciplines as superficial evidence for a structural, rule-governed model. Computational linguistics has found the systematics of turn-taking and the organisation of repair problematic if a formal specification for generation or categorisation is required. Again this is missing the point. CA findings are not woolly because they do not specify the motivations or mechanisms for producing conduct. That is not the aim of CA. The rules or structures found by CA are predicated on the situated practices of engaging in conversation - they do not specify courses of action. They are used by participants to interactively construct the sense of their actions as accountably engaged

in conversation. There is a very real danger of reifying the findings and notions of CA. This is similar to the AI conception of 'plan', which regards action as the execution of a program. The regularities found in conversational action by CA, eg. the adjacency pair and insertion sequence, are very easily made into rules for classifying or generating. However, they are indeterminate, indexical and constituted within the practice of dialogue.

3.3 An Interactional Perspective

Conversation analysis provides a perspective on routine communicative action that is a valid alternative to the cognitive theorising of computer models. It has been shown that talk is not the exchange of messages between individual cognitive processors in which meaning is grounded in a shared background knowledge. Rather, participants interactively and methodically construct the sense and shape of their conversational activities in situ. However, the term 'interaction analysis' is used for the rest of the thesis as CA has taken conversation as its prime subject matter. This is done for two other reasons, too. First, it is not the aim of the thesis to promote the conversational metaphor as a valid direction for design or modelling. Instead, an understanding of the fundamental practices of achieving coordination and intelligibility in any interactive dialogue activity is essential for the novel design of artifacts. Second, concentrating on a conversational context may miss substantial orders of detail that are unavailable to such an analysis, though reference to conversation is essential as it is the most general and least specific language-based dialogue activity. The primary focus here is on the interactional achievement of multi-party activities involving talk, body and artifacts interleaved in time-space in which the production and interpretation of those activities is a situated achievement. Dialogue is a collection of practices in which participants are mutually engaged in coordinating their communicative actions and achieving shared understanding. Cognition is relevant but an account of practical activities must include interactivity as a constitutive domain.

From ethnomethodology, an interactional perspective must be empirical and elucidate

> *descriptions of the 'architecture' of interaction: the structural organisations through which interactants collaborate and/or compete in the construction of conjoint courses of action.* (Heritage 1984, p 18)

The intention of Chapter 6 is to locate some of the findings of the empirical interaction analysis in terms of the design of dialogue artifacts. The following quotation from the domain of improvised music indicates that the metaphor of 'ensemble' is more appropriate than that of 'interchange' for future conceptions of artifacts designed as partners.

> *8KN - (B-12)/R10 has been designed to provide a challenging context for expanded participation. Each instrumentalist in the work can contribute to the*

actual moment construction of its creative material as well as its infrastructure materials. (Anthony Braxton, Composition Notebook - Vol. 1)

One of the first to entertain a CA perspective on artifact design can be found in Schegloff (1980). He briefly summarises CA's findings about the nature of talk and gives a brief indication of the extent of the problems to be faced if a conversing machine is to be built. McTear (1985, p. 121) presents a more developed consideration of the applicability of CA methods and findings to human-computer interaction and AI. He claims there are hints of the possibility of developing more formal representations of "the rules which underlie the organisation of conversation." To substantiate this from within the field of CA he quotes Schegloff et al (1978, p. 362, fn.5): "As in the organisation of turn-taking, the gross facts which characterise large amounts of conversational data are the product of rules, and systems of rules, which operate on particular sequential environments." Sacks (1984, p. 26) similarly states that the aim of CA is to "transform... our view of 'what happened', from a matter of a particular interaction done by particular people, to a matter of interaction as products of a machinery." Also, McTear discusses the differences and similarities between the organisation of human-computer interaction and human conversation, pointing out how the findings of CA can help inform the design of 'gracefully-degrading' language-based systems that inevitably breakdown in use, and that CA could be very fruitful in understanding the nature of human-computer interaction (HCI) through empirical analysis of protocols of simulated and real HCI. Also, McIlvenny (1985) suggests that the methods and findings of CA should be useful for the design and evaluation of advice-giving dialogue systems, based on the results of a pilot study of advice in restricted circumstances using CA methods and an implementation of a simple dialogue interface based on the findings of CA.

Research has begun in a number of areas concerning the same issues. As mentioned earlier, Suchman (1987) investigated participants' use of a photocopier expert system as they tried complete a number of tasks. Attention is paid to the user's talk and sequential details of the use of the machine which are in the style of conversation analysis. The COSMOS (1988) project contains research that has used insights in conversation analysis to understand the fundamental problems occurring for participants using asynchronous computer-mediated communications, particularly conferencing or e-mail facilities. In addition, Frohlich & Luff (1989) report the design of an advice system that incorporates findings from CA in the design of the opportunities for machine and user participation in an interaction between the general public and an expert advisor in the domain of social security; Agre & Chapman (1989) have applied the insights of Suchman, with regard to the nature of plans, to the design of automata that do not plan their actions in their practical activity; and Cooper (1989) develops an ethnomethodological approach to the contextual and culturally embedded work of using a standard graphics-based microcomputer.

3.3.1 Empiricism, Situated Dialogue and Artifacts

Three issues were put forward at the beginning of this chapter, derived from the critique of computer models of dialogue, that needed clarification in order to provide a basis for attacking the problematic conception of dialogue in models and artifacts. The chapter has developed approaches to these issues that are summarised below.

First, a perspective on communicative action and practice for models and artifacts is provided by social science in the form of ethnomethodology. It delivers a foundation for explaining some of the problems that artificial intelligence and cognitive science has in dealing with practical action and activity. It is not denied that a vocabulary of plan, intention or motive is a part of everyday praxis; on the contrary. It can be argued, however, that vocabularies are not generative of nor underlying explanations of conduct - just emergent from and useful in action in an unscheduled world of multiple constraints and indeterminacies.

> *A basic research goal for studies of situated action... is to explicate the relationship between structures of action and the resources and constraints afforded by physical and social circumstances.* (Suchman 1987, p. 179)

Unfortunately for models of dialogue, the tendency has been to over-extrapolate the relation of this vocabulary to action itself and develop a logic of planned action and communication. For 'intelligent' artifact design this has meant the unappealing prospect of determining the plans, beliefs, etc. of a human user 'in the moment' from highly limited and predetermined resources. An alternative, rich source of constraints or possibilities becomes apparent if one only looks at situated action itself, giving an alternative vocabulary and method that can guide the design of interactionally dynamic artifacts and entities.

Second, a perspective on the dialogue artifact is supplied both by ethnomethodology and hermeneutics. It is not an intelligent agent that conducts conversations with people, but a social, reactive object produced for human consumption - an artifact - that is constituted or grounded through the metaphor of dialogue, and thus can be called interactive. It can be called language-using, distinguishing it from texts, because of the properties of language form in reaction, lack of a visible author and complexity of behaviour.

Third, an empirical methodology is derived from CA. It studies human conversational talk focusing attention on the situated details of human, and by extension to human-computer dialogue. The careful observation, collection, recording and analysis of the details of interaction that can be seen to be important in the achievement of that activity for the participants themselves is elaborated in Chapter 4.

3.3.2 Aims, Methods and Tools

The aim here is to explore the dynamics of communicative 'dialogue' interaction by means of interaction analysis. A study of dialogue 'through' machines was conducted to show that Grosz & Sidner's (1986) theory is unavoidably misrepresenting crucial aspects of dialogue; it was similar to the study which was the source of their own data. The possibility of using these findings to enhance computer models is discussed later in Chapter 7 but this is not a primary goal of the thesis. The analysis of the achievement of computer-mediated dialogue provides material for recommendations about the design and evaluation of dialogue artifacts. This is because the investigation explores the use of data collection and transcription methods for complex dialogue spaces mediated by machine that are similar to envisaged human-machine interfaces. The methods of interaction analysis applicable to the study are derived from CA with the aim of examining instances of conduct carefully to determine their local, methodical production by the participants themselves. The collection of repeatably observable records of events or conduct with attention to emergent sequential and interactional characteristics of the achieved dialogue was undertaken. The main aim of the investigation was to clarify the notion of interactivity in communicative interaction and to pin down and describe some of the dynamics of embodied human communicative action. The basic tools required for the study are the camera, microphone, monitor and the recorder. Video and audio recording are immensely useful for an observer's analysis of accountable conduct that otherwise cannot be studied because of the temporally detailed, 'tacit' and 'unrecognised' nature of that conduct. Also, a novel notation system has been developed for vocal, gesture and gaze behaviour that is suitable for locating the work of using a computer-mediated communication device.

Chapter 4

AN AUDIO-VISUAL STUDY OF COMPUTER-MEDIATED DIALOGUE

4.1 Introduction

This chapter introduces an empirical perspective. An investigation of dialogue 'through' the machine - or computer-mediated communication (CMC) - was conducted, and there follows a broad introduction to the study in order to map out its basic constituents. From the arguments in the last chapter, what is needed is an empirical and descriptive approach to the study of human action and discourse. In the rest of the thesis, such a methodology will be used to reveal some of the contingent and situated aspects of human dialogic interaction, and to give a perspective on machines that are to take part in activities such as communication and interaction with human users.

The two basic points made in Chapter 2 were as follows. First, it was argued that computational approaches to dialogue take a particularly weak view of interaction, and disregard the situated character of human conduct in order to make general and abstract claims about discourse. Second, it was suggested that if such a proposed narrow perspective on dialogue is adopted, whether to simplify the data or to focus on the design of interactive artifacts, such as natural language interfaces, then crucial features are ignored, which leads to an abstract and ad hoc theory of dialogue and interaction. So, to counter this tendency an empirical study of human dialogue 'through' machines was initiated in order to take a closer look at the nature of dialogic interaction. The study examined the achievement of dialogue between people when it is mediated through the use of the computer. Most everyday interaction takes place in face-to-face encounters in which the body plays a crucial role. Other dialogues are undertaken, for example, on the telephone or in the form of correspondence or texts. However, the computer can be used as a mediating technology that allows for dialogues to take place in real-time or contexts of presence that are markedly different from conversation and correspondence. So, rather than investigate asynchronous electronic mail, which is rather similar in many respects to correspondence, we will examine

a synchronous CMC that allows people who are spatially but not temporally distant to take part in dialogue even though they may be a hundred miles from each other. The particular CMC used is a relatively old example of collaborative technology that is widespread in a similar form on many multiuser mainframe computer systems supporting a large community. It allows spatially separate users who are connected by the same computer to construct typed 'messages' on a shared virtual screen supported by each monitor, ie. I see what you do and vice versa or 'what you see is what I see' (WYSIWIS).

The initial motivation for the study was to look at a CMC that is similar to the dialogue modalities envisaged as feasible or suitable in human-computer interaction (HCI) and artificial intelligence (AI). McTear (1987) claims that the study of modalities of communication and the simulating of interfaces are useful research directions for AI and HCI that can also inform the study of human conduct. In this case, Grosz & Sidner's empirical methodology is shown to be deficient, and the situated, embodied local achievement of dialogue turns out to be crucially important. For unfamiliar and odd dialogue modalities that are bound to develop unintuitive interactional dynamics in situ, Grosz & Sidner's model is hard to apply because almost no account is taken of circumstantial and interactional features of the production of dialogue. Our central interest is in how dialogue is achieved between participants who can engage in synchronous mutual dialogue activity within the constraints of the modality. Since the complexities of everyday conversation have been documented in different ways by conversation analysis, it seemed more useful to investigate communication between human beings 'through' the computer. This particular mode of communication is introduced because it is generally available for human user consumption and is used by some sectors of the computer-using community. Moreover, it has its own peculiar emergent dynamics and characteristics - shown in pilot observations and recordings - and the disruptions caused by a lack of resources or the strange circumstances highlight routine features of interactivity, since the disturbances and breakdowns to the normal flow require local mutual adjustments on the part of the participants in order to achieve 'sense'.

In parallel, an investigation of copresent collaboration is developed. The motivation for this step that is documented in Chapter 6 is to draw upon two contrasting communicative activities: the visual-based virtual CMC dialogue and the copresent dialogue over the workspace. Because technology, representational devices and artifacts are an integral part of our everyday practical life and work, an understanding of talk not just as conversation but endogenous to practical activities is essential. Talk takes place in the context of other activities and practices, and is organised in and through those activities as an integral part of them. In addition, as will be argued in Chapter 6, the primacy of the model of copresent conversation for the design of dialogue artifacts and spaces must be challenged, and this

challenge requires us to understand how interaction, dialogue and embodied activity shape each other.

For various reasons the study was conducted in an experimental context that diverges from the primary work in conversation analysis. Participants were instructed in a simple task requiring everyday practical reasoning in order to plan routes with the help of maps. For successful orientation, the participants had to cooperate through dialogue. In order to gain access to some of the work in achieving dialogue and the task, a technique for the elicitation of collaborative activity - not only verbal reports - was used. Also known as 'constructive interaction', the technique requires that two co-participants collaborate in an activity they could both do on their own. In the course of engaging in the activity they display, in their actions and talk, interesting accounts of their own and the other's behaviour, which is available to an observer or analyst as well. Specifically, two subjects are required to communicate through the modality with two other subjects, and thus each pair display accounts in their talk of the sense of the dialogue in progress between the pairs. In addition, the study investigates whether CA methods can be applied to an unusual communicative activity. Techniques of observation and description had to be developed, including video and audio recording and a type of transcription that would record the participant's 'observable' practices, making them available for repeated inspection. Features of vocal, gestural, gaze and CMC activity were transcribed using a set of notation conventions derived from CA, but a quite different and complex record structure was used in which streams of conduct are rendered in relation to the CMC activity. As the modality has features that resemble the human-computer environment, transfer of interactional analytic methods to that domain is a possibility entertained in Chapter 6.

4.2 Related Research

Before describing in detail the principles and procedures used in the study, a survey of related research is necessary. This will be grouped under headings that reflect the different aspects of the study: modalities of participation and interaction and their effects on communication; similar studies of computer-mediated communication; constructive interaction techniques; instruction and task-oriented activities; maps as representational resources, and route-finding.

4.2.1 Modality

The term 'modality' or 'mode' when applied to communication is a description of the constraints, conventions and resources that participants have at their disposal and use to construct their communicative activity. This is different from the concept of 'medium' which

is traditionally used to describe the qualities of the expression of communication, eg. speech. The modalities of writing and talking have long since been recognised as distinctly different modalities of communicative activity, but the range of distinctive modalities besides these two and how they are achieved has only been of recent interest. Several studies of modalities similar to the CMC used in the study have been reported. Reflections on the relative power of writing and talking have been the preoccupation of generations but this survey will only go back as far as studies of modalities that relate to computer-mediated communication.

Chapanis et al (1972, 1977) report a broad psychological study of problem solving among subjects communicating through a restricted modality in order to solve the set problem. The modalities studied were a simple Teletype interchange, handwriting, voice, and face-to-face. Statistical measures of time to complete a problem, number of words used, etc. were the basis for conclusions about the gross effects of the modality on behaviour and language. Also, conclusions were made about the relative performance of the modalities for solving the tasks. Chapanis (1981) adds other modalities: the telepen shared graphic screen for drawing and writing, and the video link. Also, the influence of technology on communication practices was now explicitly recognised in the statement about discovering "how interactive human communication is affected by the machine devices and systems through which people converse." (p. 66) However, in all the studies only simple records or reduced scripts of the communicative problem solving activity through the modality were kept. Only the products of a communicative activity were used as evidence, so the circumstantial production of that activity is lost. For example, only the Teletype output was kept, or a word-based script of the telephone dialogue was transcribed much like a text of an interview for a magazine article. Similar practices have been shown in Chapter 3 with respect to Grosz & Sidner's empirical methodology.

Burke (1982) reports a similar study that looks at face-to-face, telephone, audio-letter, and handwritten message modalities used by subjects to cooperatively solve a problem. He takes a more promising 'constructive-interactional perspective' in which communication is an interactional process of coordination organised by communication intentions. However, the theory is weakened by the lack of attention to the empirical details, a reliance on intention, and the quantitative statistical analysis of coded transcripts. In spite of this, he does note an interesting phenomenon also documented later in this thesis, namely 'self-transcription'. That is, participants in the handwriting mode must render their thoughts and construct a communicative message using graphic symbols. This is in some ways a similar activity to what the transcriptionist is engaged in when rendering the speech of conversants. Grosz (1977) also reports experiments involving restricted modalities conducted as part of a research project. Grosz & Sidner's task dialogue example discussed in Chapter 2 was drawn

from this corpus of Grosz's, and the chapter described the odd restricted modalities in order to demonstrate the weakness of their data. Basically, the dialogue between the apprentice and the expert was really mediated through another 'message-passing' human intermediary, so they did not have a common modality: the apprentice communicated using the telephone modality and the expert using the Teletype modality. Additionally, for some dialogues the participants had limited mutual visual access because the expert could request snap-shot pictures of the problem-solving space.

Cohen (1984) brings together the two strands of work that focus on task-oriented dialogue and modalities to look at the pragmatics of reference. The modalities reported to have been empirically collected are the non-interactive written and audiotape, and the interactive face-to-face, telephone and keyboard, though the work reported in the paper concentrates on the latter two. However, the exact nature of the keyboard mode is ambiguous: the comment that "simultaneous typing was possible and did occur" and the "subjects were informed that their typing would not appear simultaneously on either terminal" obscures what is going on. The analysis invokesa traditional plan-based speech act approach to generating the coded acts - basically requests and informs - that have strict theoretical definitions. Unfortunately Cohen does not consider the full circumstances of communication to be important for his concerns, so he only records a 'script' - a residue of what was typed - of the keyboard interaction, and transcribes simply the utterances or 'word and sentence speech' of the telephone dialogues. There are a number of insightful comments though. He warns that the results and theory from this particular experiment should not be generalised to human talk and discourse for the following reasons:

* The task studied means that only physically present objects are talked about.
* The telephone and keyboard modalities do not accurately reflect all natural language communication between a person and a machine.
* It is difficult to transfer results from conversation between humans to HCI interface - for example, indirection may not be common in future speech-language systems as it is in human talk.

He notes in conclusion that, although most corpora in computational linguistics have been of dialogues conducted in keyboard modalities, his experiments demonstrate that keyboard communication is distinctly different in structure from telephone and written communication. "We should thus be wary about theories' and techniques' coverage if they are to [be] extended to other modalities of communication." (p. 123)

Oviatt & Cohen (1989), looking closely at the differences between spoken communication and written modalities, renew Cohen's warning about which models of dialogue are

appropriate for the design of computer dialogue systems. Spoken dialogue is delivered more rapidly, tends to be less planned, less concise, less complex, less well integrated syntactically, fewer abstract ideas, shorter and less varied vocabulary, more pauses and disfluencies, more hedges, quantifiers and function words, more self-reference and pronouns, more requests for confirmation, more repetition, more noun phrase reductions with repeated reference, more indirection, a more fine-grained decomposition of requests, and more metacomments (Oviatt 1988, p. 44). Also, an important distinction among speech modalities is interactivity. By noting the differences that manifest themselves because of the interactive nature of most speech they argue that spoken monologue should be taken as the model for building the next stage of speech dialogue artifacts. This is a commendable effort to refocus design on the problems envisaged because of interactive circumstances but diverts attention to the speaker and the abstract formulation of speech that has been argued against in Chapters 2 and 3.

All of the empirical studies of modalities of communication have focused on the traditional notion of communication between the participants, ie. they regard adequate data to be a sedimented script or record of certain aspects that the analysts interpret as being communicatively relevant and available to both participants. So, parcels of words and decontextualised message units are the basic units of communication that must be documented, if at all. The study in this thesis aims to show that this focus loses the situated character of production and interpretation and leads to a weak and one-sided view of interaction and dialogue. For example, an interesting study is reported in Suchman (1988) that concerns face-to-face whiteboard practice - a large white board on which people can write and draw with coloured pens - in research settings. Unlike the modalities described above, the whiteboard is not a distinctive 'talk' modality but an extra, parallel, optional space for writing and drawing relevant to the scientific activity in progress. However, this study is relevant as ethnomethodological techniques are used. Audio-visual recordings are made of work on a whiteboard and the analysis focuses on "the organisation of activities that produce marks on the whiteboard and gave them their significance, and the function of the marks in the structure of the activity." (p. 318) In this case, Suchman has the distinct advantage of being able to analyse the talk that emerged directly from the joint work of the participants on the whiteboard. Thus the abstract 'script' or record of the joint activity on the whiteboard - the 'mess' on the board at the end of a session - can be analysed in terms of its interactional production.

4.2.2 Computer-Mediated Communication

No studies of the same computer-mediated communication[1] modality using a similar method to that reported here have been found. But studies of different aspects of similar synchronous modalities are available. The closest is that of Daly et al (1989), who report an experimental investigation of communication between two subjects using a similar modality to strike up a casual conversation. Another study, Kiesler et al (1985), has investigated affect, from a social-psychological perspective, in a similar modality comparing it with face-to-face communication in an experimental setting. This study is rather different from others in that quantitative measures are taken of subjective affect and expressive behaviour in order to test the hypothesis that restricted modalities depersonalise communication. Again, as in the last section, only scripts of the computer-mediated dialogue were retained for inspection, and additionally they were reformatted to make them indistinguishable from the verbal transcripts of the face-to-face talk so that they could be used in a further "unbiased" coding stage. Also, Kennedy et al (1988) report an experimental study of a simulated expert system that uses a similar but more restricted modality for communications between the subject and the simulator.

Studies have also been made of asynchronous modalities like electronic mail, which is in prolific use in the academic community. Black et al (1983) note the interesting phenomenon of multiple threads of topic in electronic mail that is similar in nature to the double dialogues in the CMC modality analysed in Chapter 5. For instance, regularly in the exchange of electronic mail messages participants develop multiple topics in the same message that are available for responding to in the reply, ie. much like letter correspondence. A brief look at scripts of authentic dialogues between users and operators of a synchronous CMC showed that such a phenomenon is not routine in more interactive real-time dialogues, though the development of multiple topics over a period of time is normal. This point does, however, indicate the need to rethink the common assumption of strict adjacency in dialogue: that is, the idea that participants construct dialogue as a linear stream in which a turn sequentially implies only one adjacent response and only addresses one adjacent aspect of the prior dialogue.

The only study taking a conversation analytic perspective on the nature of computer-mediated dialogue is that reported in COSMOS (1988) - a report of a project on asynchronous message

1 Bannon (1986) contains a discussion of the different forms of computer-mediated communication.

passing systems, eg. for computer conferencing. The concept of adjacency in CA and an analysis of the temporal dynamics of message transmission, reception and response are used to demonstrate the faults with a speech act model, and the peculiar properties of electronic mail that lead to breakdown.

4.2.3 Constructive Interaction: An Analyst's Resource

The methodology for the audio-visual study reported in the thesis uses the accounting practices of talk as an analytic resource for locating just how communication is achieved in the modality. Two people constructively collaborate on achieving intelligible communication through the computer with another person and thus display to each other in their talk and activities the sense of the ongoing dialogue. The method closest to that used here is found in Miyake (1986) who used 'constructive interaction' (a method of eliciting talk between participants collaborating on a task) to locate how people come to understand a complex physical device, eg. a sewing machine. Audio-visual techniques were used to record two subjects talking in activity. O'Malley et al (1985), drawing on Miyake's work, also recommend it for evaluation in human-computer interaction.

Daly et al (1989) report an experimental psychological study of cognition in computer-mediated communication dialogues between strangers. They used a protocol generation technique in which subjects talk aloud while engaged in their activity. (See Ericsson & Simon (1984) on verbal reports as data in psychological research.) Video and audio recordings were made because they recognised the importance for the analysis of placing a given comment in relation to the ongoing dialogue activity. However, the method is subject to the objections that the participants' reflective comments on their activities prompted by the experimenter bear no necessary relation to the internal accountability of the activities (nor tacit internal processes), and the presence of the experimenter is an external interference. Moreover, the achievement of communication in and through the modality is taken for granted, whereas in this thesis the achievement is made the object of study.

The work of Suchman has already been described in Chapter 3, but her empirical methodology has not been explained. Let us consider the following example of her approach to understanding the use of intelligent artifacts such as expert systems. She studied people using a photocopier help system in which a rule-based 'expert-system' with a simple textual and graphic screen display was triggered by certain physical actions of the user or machine. The initial aim was to find out why naive users found the help system confusing, when it seemed quite straightforward to the designers. Videos were made of pairs of first-time users using the photocopier to complete some tasks predesigned on the basis of earlier pilot studies. A corpus of four sessions each lasting one and a half to two hours was collected

and transcribed, to show the differential access that the machine and the user have to the circumstances of interaction. By using two users, 'constructive interaction' protocols are generated that expose some of the work done in order to complete the tasks set by using the interface. Video technology was used to capture some of the bodily work of collaboration and device use. Suchman's research is one of the first to apply the insights of ethnomethodology and CA to the interface.

4.2.4 Task Dialogues and Instruction

Because of the experimental context of the study some motivation was needed for participants to engage in dialogue through the computer. A task was chosen that required participants to cooperatively construct a shared route through a city using a map as a resource. Task-oriented dialogues have been the focus for a number of investigations. In Chapanis (1972, 1977), subjects had to solve one of two problems. One was a geographical orientation problem involving cooperation in finding an address given that a map had been given to one participant and a telephone number book to the other. The other was an equipment assembly problem in which one participant is given the parts to be assembled and the other the instructions; together they must cooperatively communicate in order to construct the object. Deutsch (1974) and Grosz (1977) report the work around the SRI project on the toolkit workstation that looked at task-oriented dialogues. The task under study - the toy water pump assembly - became a standard for many future studies, such as Burke (1982) and Cohen (1984). In Burke(1982), the participants had again to cooperate in constructing the toy water pump, but the problem-solving tasks were variations on a theme: a single subject with no instructions; two subjects collaborating with no instructions; two subjects with only assembly-sequence cues; two subjects with only spatial cues.

Not mentioned so far is Bateman's work on discourse resources for achieving mutual intelligibility in a game (Bateman 1986). It is based on the maze game reported in Anderson (1983) and Anderson & Garrod (1987). The design of the computer-generated game was such that cooperation and mutual orientation were essential for the successful completion of the game. Participants were in separate rooms with only an audio link with which to communicate about what moves to make in the game. The display was designed such that information necessary for moving in the maze was only available from the other subject, but with an added twist - a move would effect the possibilities for the other to move. Audio recordings were made and game states automatically recorded by the computer. The main relevance of his work is to look at linguistic form and discourse functions in the achievement and management of a 'world-in-common' for all practical purposes.

4.2.5 Maps and Direction-Giving

Because it was decided that representational maps were to be used as the interpretative resource for constructing a shared planned route, the work of ethnomethodologists on such practical matters should be noted. Psathas has studied direction maps and direction-giving in a variety of settings. The earliest paper, Psathas & Kozloff (1974), looked at the structure of directions in telephone dialogues given in order that the caller could get to the other's location. They are interested in what the basic elements of directions are that give them their sense, ie. how are they seen to be adequate. The work in Psathas (1986a, 1986b) continues looking at telephone directions, but this time, in terms of their sequential organisation. His main finding is that the sequential nature of talk is a resource for organising the sequence of operations seen as describing movement to a destination, ie. directions to a place are a local interactive achievement through talk. His interest in these papers is in how possible relationships in space and time can be interpreted from the interaction, rather than the correspondence between the relationships described and those in the 'real world'. In a slightly different vein, similar to Suchman (1988), Psathas (1979) analyses graphic direction maps written for an absent other so that they may find their way to a particular place familiar to the author - namely, home. Here, the question is: how do the practically organised set of features produce the sensible quality of their being 'maps' for locating a place? The emphasis is on the direction map as a potentially usable map, not in how it is used to actually get to the place. They can be examined "to discover the methods of practical reasoning which their users depend upon to make what they say understandable." (ibid, p. 203) In the study here, participants must construct sensible routes through a city with only a graphic representation of that city. Because the participants have different maps, discrepancies in their details expose the maps as contingent representations of a physical city space that require local and interactional interpretation. However, an investigation of how the sense of the maps is locally achieved is not the focus of the thesis.

4.3 Principles

> *The work is part of a program of work undertaken... to explore the possibility of achieving a naturalistic observational discipline that could deal with the details of social action(s) rigourously, empirically, and formally.... Our analysis has sought to explicate the ways in which the materials (records of natural conversations) are produced by members in orderly ways that exhibit their orderliness and have their orderliness appreciated and used, and have that appreciation displayed and treated as the basis for subsequent action.* (Schegloff & Sacks 1973, p. 289)

As the thesis reports an interactional perspective on computer mediation, models and artifacts, derived from conversation analysis, and as CA takes a primarily naturalistic empirical stance, this section will describe the principles underlying the empirical study conducted. Before this can be done, however, it must be noted that even though there are a range of procedures available from CA and ethnography, it has not as yet been established whether principles of the study of everyday human talk may transfer generally and successfully to the study of machine-based transformations of dialogue activity[2]. One principle of work in ethnomethodology is to investigate the intrinsic taken-for-granted character of a practice, but Heritage (1984, p. 302) notes that "there are scarcely any straightforward methodological pathways which presently inform the descriptive enterprise." Since there are no stock methods which will apply to an investigation of a particular activity, the study - finding out how intelligible dialogue was achieved as a practical accomplishment by those engaged - borrowed some basic principles from the study of conversational talk, but also developed methods appropriate for the task in hand that deviated from CA practice.

First of all let us look at the similarities to and differences from standard CA practice. Livingston (1987, p. 10) has stated that ethnomethodology is interested in instances of naturally organised everyday activities, ie. that "the activities under investigation are ordinary, that they are organised, and that organisation is natural in the sense that it is part and parcel of the activity itself." This perspective has distinctively defined the field and focused its research to exclusive domains that are penetrable without flouting this principle. In particular, CA has mainly studied everyday casual conversation in marked contrast to other fields studying language use. Heritage & Atkinson (1984, pp. 2-5) illustrate what the empirical stance entails by considering the traditional alternatives. These include the use of interviews and questionnaires in order to discover the participants' views of the conduct they have been engaged in. Such interviews and questionnaires are excluded, because they bear no accurate relation to actual conduct. Participation in an interaction by an experimenter, who may take notes, is also excluded as a means of gaining access to the interpretative practices of a group. To gain records or evidence of language use CA particularly excludes introspection and made up examples. Also excluded is apparently unbiased coding by the analyst or by a paid coder using pre-established coding schemes, because the latter may exclude interpretations that may develop during analysis and the schemes are almost impossible to apply consistently. All of the above are excluded from the study here so as to

2 Nor has it been established up to now whether this is also the case for analysing user's activity in and through dialogue artifacts.

analyse the natural history of observable phenomena by a competent member researcher: this is a foundational principle of CA research.

One further consequence of the principle is that experimental settings are considered inappropriate when the aim is to investigate without bias naturally occurring conduct in real practical settings. However, an experimental setting is used in the empirical study in the sense that the task is constructed and subjects are elicited to take part. The aim of the study is to explore the issues that arise in the course of researching modalities of communication, when it is difficult to study 'natural' use of that modality. What it means is that one has to be very cautious about using the methodology and claims of CA or ethnomethodology, especially as they have always been very careful about the authority and generality of their claims in relation to their 'objects' of study. But, Garfinkel (1967) reports studies of the normative nature of human conduct that use experimental methods to elicit evidence for the documentary methods, eg. subjects were required to take part in an advisory encounter in which only 'yes' or 'no' answers were given to the subjects' requests for advice. In addition, a protocol generation technique was used to elicit the subject's interpretative orientation to the advisor's responses. Also, Suchman's study of the situated use of a plan-based photocopy help system was carried out in an experimental context. Observations led to the design of a set of tasks to be carried out by pairs of subjects as if they were naturally occurring (Suchman 1987).

It may have been possible to investigate computer-mediated facilities, eg. electronic mail or the same CMC, in everyday practice. But co-participant protocols, which give some analytic access to the chronic process of production of the interaction or interchange, are not usually an integral part of the everyday performance. Also, for the CMC studied, it would be difficult to access the temporal flow of communication without disturbing the setting and interaction. For example, to trace and record the physical events in the interface that are significant for the participants or for the machine requires tampering with the resident equipment or software that may effect the operation of the equipment or damage it. Equally, to video the encounter, if it is appropriate to use such a technique at all, may be impossible because of distributed, non-localised, and unpredictable engagements[3]. Instead the study developed as a sort of uncontrolled investigation in which specific objectives were not designed so that as many

3 The study is rather different from one that might intend to video telephone users in their separate locations to determine if their actions 'off stage' were immediately relevant to their mutual conversation on the telephone. Here the conduct 'off stage' from the CMC is rich and crucial for the achievement of dialogue in the CMC modality. It is not immediately relevant as such to the mutual dialogue, but it displays the work of constructing and interpreting the developing dialogue for that party.

lines of interest were kept open. Different types of modality data were collected and as much of the events recorded as possible in order to extract or 'find' interesting material during the transcription and analysis. Also, after clarification of the task instructions the experimenter did not tamper or interfere with the setting.

The study excludes from its methodology most of what CA regards as problematic in the study of human social conduct, but does take an experimental perspective in order to examine closely the activities of communication in the CMC. Of course experimental effects were apparent. This was shown in the subject's orientation to environmental events: ignoring of telephone ringing, noticeboard distraction, and noticing of camera. However, the same is true for recording of naturally occurring events, eg. Heath (1986, p. 11-13) presents an analysis of a patient's orientation to the camera.

Another principle is to found an analysis on 'emic' data, ie. phenomena, resources, understandings, etc. that are an implicit part of the participant's 'work' in producing the conduct. That is, they are not only an insider's interpretation or recollection after the event, but endogenous to the conduct being investigated[4]. To expose some of the work done by participants in interpreting and making sense of the interaction in the CMC, a method of protocol generation is supported. The generation of verbal accounts is clearly an integral part of some settings - where the need for talk in the setting is essential for the achievement of the circumstantial work. However, in this case not only is the setting artificial in two senses - being neither a 'real' issue nor necessarily a 'real' task that would be routinely accomplished in the ways outlined - but in addition the number of participants is determined by the purposes of the experiment. Basically, two co-participants do the work of the solitary communicator. This situation is not normal, but on the other hand it is not untypical, as people sometimes do this in real situations of use. This method has two aims: the talk of coparticipants reveals constitutive practices as they work on the task in and through the CMC modality; and the collaborative activity of the coparticipants reveals rich and complex dimensions of interactivity in copresent problem-solving.

One essential source of 'emic' data, an analyst's resource, that has informed CA work and is still available in this modality is the displays of adjacent interpretation both in the CMC and coparticipants talk:

> *...just as a second speaker's analysis and treatment of the prior is available to the first speaker, so it is also available to overhearers of the talk, including social scientists. The latter may thus proceed to analyse turns at talk, together with*

4 Within linguistics research the concept of 'emic' data is contrasted with 'etic' data that denotes the extrinsic analyst's constructions that are not available to the participants in that conduct.

> *the analyses and treatments of them that are produced by the parties to the talk, and employ methodologies that fully take account of these analyses and treatments. Students of talk are thus provided with a considerable advantage that is unavailable to analysts of isolated sentences or other 'text' materials that cannot be analysed without hypothesizing or speculating about the possible ways in which utterances, sentences, or texts might be interpreted.* (Heritage & Atkinson 1984, p. 9)

However, there is a bias in using this resource that excludes consideration of the displays that a piece of talk activity gets while in progress. Goodwin & Goodwin (1987, p. 4) argue that "the treatment that a bit of talk gets in a next utterance may be quite different from the way in which it was a heard and dealt with as it was spoken." This is a demonstrably pervasive phenomenon in the CMC because of the extended temporal development of turns that allows treatment of the 'talk' in progress. Also, this point will be taken up in Chapter 6 concerning the achievement of simultaneous actions and concurrent activities.

Another principle that is crucial to the descriptive enterprise is that the details of interaction cannot be disregarded or taken for granted; an analyst must have an open mind (Heritage 1989). In the study, multiple and interleaved streams of conduct are relevant in simultaneous and reciprocal activities, eg. vocal talk, body and gesture, artifact use, and map reading. In order to capture the 'visible' accountability of these activities, and the routine process of monitored production, audio-visual recording technologies were used. Thus, 'observable' events are made available for detailed examination and review. This method is preferable to note-taking or participant observation because of the limitations of memory and the constraints of watching and documenting multiple activities in real-time, even with multiple observers. However, recording does pose its own problems. For example, the presence of camera and associated technology is not to be disregarded as an effect on the participants' conduct. However, for ethical reasons the apparatus and the intentions of the experimenter were disclosed as subjects were informed of the purpose of the experiment and the reasons for and future uses of the recording.

A perspective on the copresent interaction of co-participants emerged from the transcription of details found in a strip of conduct. However, as mentioned above, rendering is not as straightforward as it may seem, especially when investigating a new activity. Decisions need

to be made about the means of representation that do justice to the phenomena. The CMC has important characteristics similar to those in conversational activity, eg. pseudo-speech stream, thus methods can be derived from those used in the study of conversation, though the relationship between the work in the modality and copresent activity is problematic[5]. The transcriptions were not made in the form of an abstract script of the inscriptions produced in the modality, but were under continuous and increasingly detailed development in order to be faithful to what was accountably 'real' for the co-participants and participants. Predominantly, the CA transcription notation is used for talk, so the principle of impressionistic rendering is adopted (Jefferson 1989). Additionally, a system for notating gaze and gesture is developed similar to that described in Heath (1986) and Goodwin (1981). Other aspects of the notation and records were developed independently in the context of rendering this particular activity, eg. notating the screen activity, and the graphic notation of simultaneous activities based around the temporal evolution of computer-mediated communication. The data corpus itself is unfinished; it is not coded but open for retranscription or reinterpretation because a record of transcription is not definitive. The analysis proceeded through inductive cycles of transcription, analysis and retranscription according to the emergence of interesting phenomena that needed closer inspection.

It is relevant at this point to note just what the status of the data is in the analysis. Sharrock & Anderson (1986) argue that the data is not the representations but the common-sense understandings and interpretative practices of people, and thus "objectifications in the form of tapes, etc. are useful in enabling us to become aware of and to articulate assumptions, thoughtless interpretative practices and so on." (p. 111) Because of this, Carter & Anderson (1989) warn of the unreflective analysis of audio-visual recordings that regards the detail of the recording as revealing the same order in the actual phenomena. It needs to be shown that the features identified from audio-visual analysis are relevant to the conduct, and that

5 For other types of modalities or HCI in general, how orientable events or actions should be represented and how relationships between them should be made explicit is not at all clear, and has almost no precedent in research investigations. It will be necessary to also include machine states or design motivations/descriptions. However, it is much harder to claim that the activity of the user with the machine makes available a display of interpretations when much of the user's actions are not involved in the mutual construction of intelligibility, but in solitary, silent, reflective activity.

the situated work is actually evidenced in such a record rather than created as an artifact of method. This will be undertaken in Chapter 5 with respect to the study reported here.

4.4 Procedures

In order to collect data, the experimental context had to be designed in terms of its geography, tasks, subjects and conduct of the study[6]. The decisions taken will be described with respect to the principles stated in the last section.

4.4.1 Physical Layout

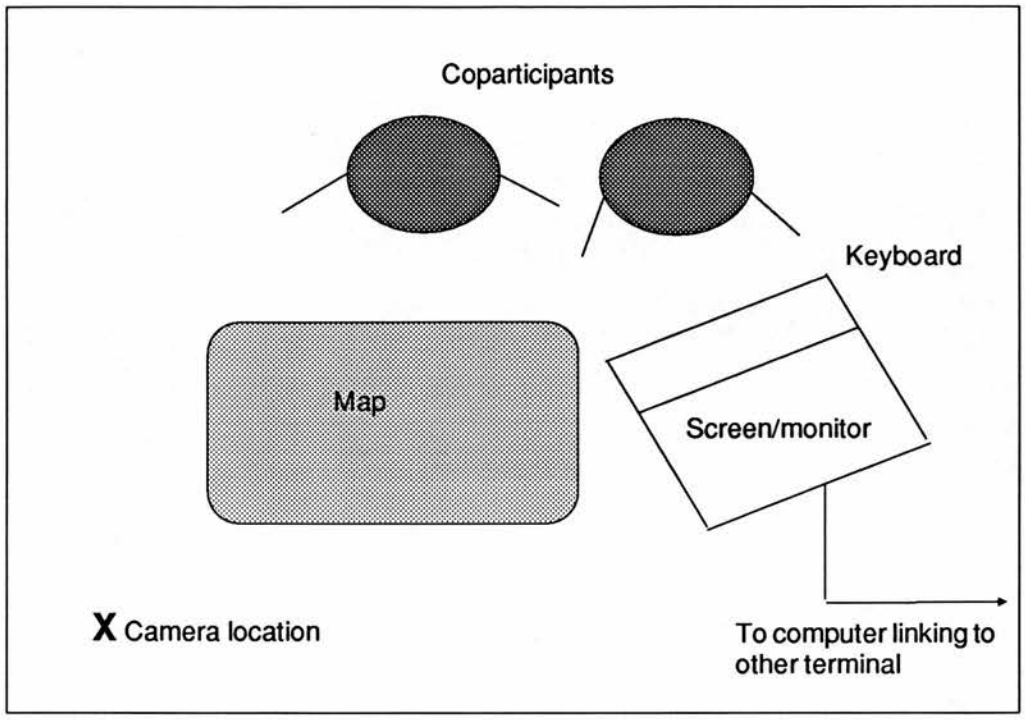
The physical geography of the experiment revolves around the limits and constraints of the technology involved. These restrictions are at present quite a handicap, so much so that naturalistic studies in real everyday environments, like workplaces, are an impossibility without interfering with the technology in use and incurring difficulties with remote and synchronous recording. The experiment was conducted in two large adjacent rooms that allowed the recording of synchronous and remote events. The recording equipment and the technician, who monitored the recordings, were located in one room behind a screen so as not to be visible to the subjects. The subjects in each room were positioned in relation to the artifacts and each other as in Figure 4-1.

They were seated and both were able to see the map and screen at the same time. A black and white video camera was located in each room in such a position as to record the participant's activities from the front, face on (See Figure 4-2.). Special lighting was used in order to obtain good contrast in the final video image. Eye gaze, body posture and gesture were all visible, but the exact location of a pointing gesture or gaze could not be determined from just visual inspection. Stand microphones were placed close - but unobtrusively - to the subjects in order to record onto the video tape the speech and noise relevant for the activity engaged in.

A complex recording set-up was developed in order to capture those details felt to be relevant for a full analysis. Two video recorders, two time-stamp generators and a split-screen mixer

6 Additional modalities that were also recorded but not used in the final analysis had the following set-ups. For the telephone mode, two participants were in the same room, separated by a screen without visual contact and they were audio-recorded. For the copresent mode, two people were sitting together, working on their own maps which are mutually visible, and they were recorded in video and audio. For the keyboard interchange mode, two terminals are linked using a standard facility that allows strict alternation of turns in which the participants are allowed complete control of their turn and the point of transfer of the floor to the other. Machine logging or trapping of temporal events was used.

FIGURE 4-1 - A plan of the physical setup of the study

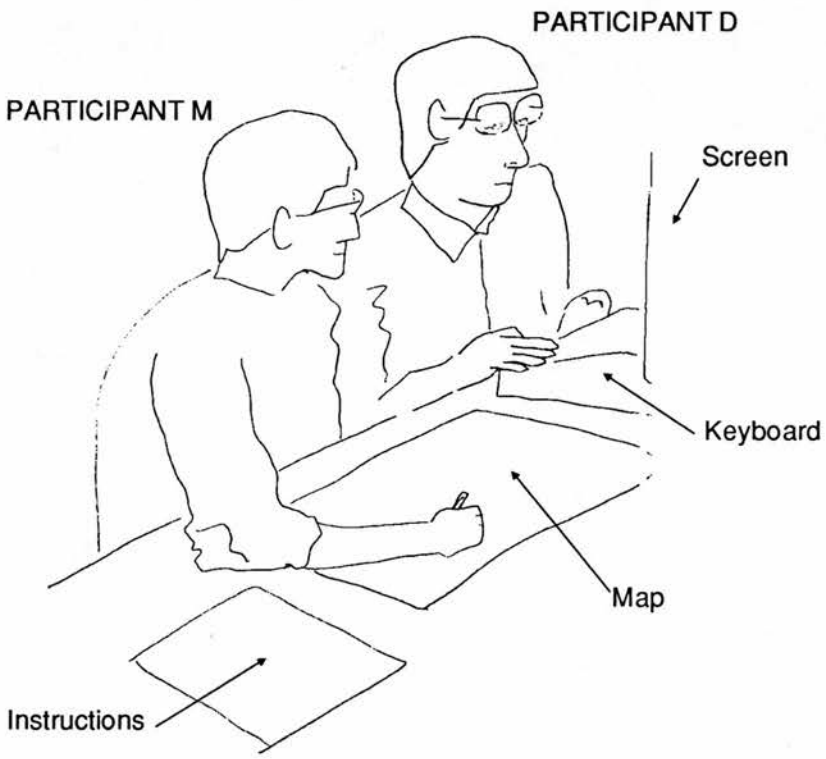


were required in order to record the details of the conduct so that the synchrony of events in time could be recovered from the video and audio recordings. The images from the two cameras were recorded on two separate video recorders. From one room - call this the subroom - the camera image was recorded onto a VHS recorder in mono audio; in the other room - the mainroom - the image was recorded on a U-MATIC recorder using a mixer to merge the camera image with an image taken from one of the monitors, which is identical for both terminals. The mixer can merge the two images into a resulting image that displays them side by side such that the visual and verbal activities of that group can be seen synchronously with the ongoing dialogue displayed on the monitor. In order to synchronise with the activities of the other room, two extra details needed to be undertaken. First, the U-MATIC records in stereo, so the audio from the subroom is also recorded in synchrony with the audio track of the mainroom, as well as being recorded on the VHS. Second, in order to demonstrate synchrony of the two video tracks, two synchronised time-stamps were used to mark a clock time on both video tracks. Both of these complications were made necessary because the mixer could not merge three images into one video image. This meant that the process of transcription to be described later in this section was complex and time-consuming.

Dialogue is established through the standard PHONE facility, a computer-mediated communication modality, on the VMS computer system that allows two remote users to communicate by typing at their keyboard. The Acorn BBC microcomputer was used to

simulate a normal terminal format for the VMS system. It has a standard layout of QWERTY keys, with some extra function keys[7], eg. a delete button plus others that were not explained to the subjects. Two virtual terminals that simulate the terminals required to use the computer were connected to the departmental mainframe VAX machine and the PHONE facility was

FIGURE 4-2 - Sketch of video picture from [M5:8'13"]



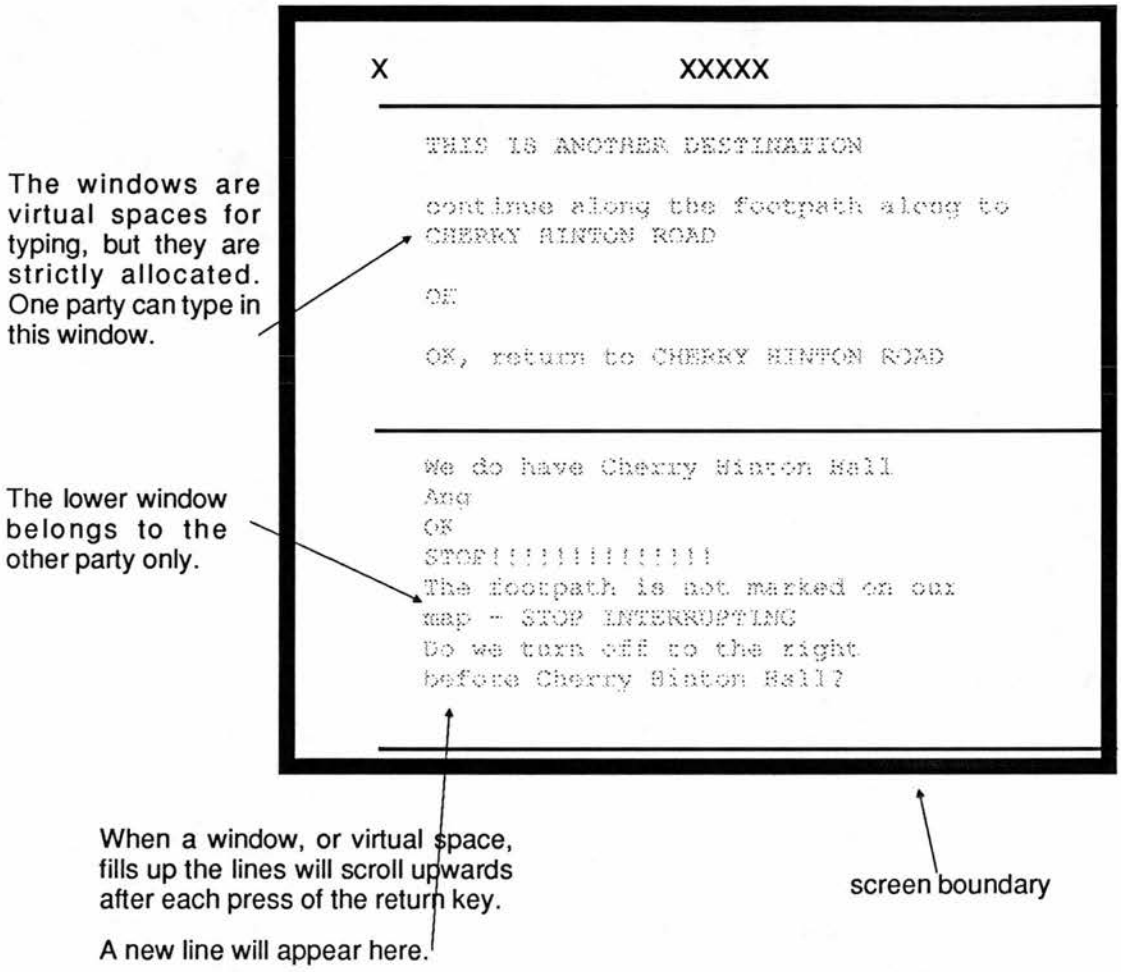
run in order to create a computer-mediated link between them. The characters typed by each user appear in their personal, visually separate windows instantly on both terminals, and both can type simultaneously. Thus, the users have no imposed turn-taking system - the floor has to be managed locally in and through the interaction itself. The diagram in Figure 4-3 illustrates the basic layout as it appears to the participants in this study.

This example illustrates the appearance of the visual communicative space from the perspective of the participants. It is taken from the later part of one experiment at a particular

7 One participant who was familiar with the BBC microcomputer keyboard did use one of its special functions in the course of the experiment.

instant, viz. [M5: at 21'26"], and records exactly what appeared on the screen. There is no analytic transcription[8]. The top half of the display screen is windowed for participants D and M, and the bottom for J and A. The two windows are identical in function. Note, the screen

FIGURE 4-3 - Visual appearance of the screen in the 'PHONE' CMC modality.



had to be reduced to half its normal horizontal width because of the limitations of the recording apparatus.

4.4.2 Task

The experiment ran as follows. Subjects were instructed that they were taking part in an experiment involving one of several communication modalities and that they would be asked

8 Only the result of the participants' self-transcription when constructing practical typed responses is displayed in deliberately faded text.

to cooperate with the other party in completing a task. The basic aim of the artificial task was for the parties to plan a route through a city using a copy of a published "city map". One party is given a map which has some destinations that must be visited, and some routes already marked that must be followed. This party is instructed to find a route that visits all the destinations and follows all the routes, and to guide the other to do the same. The other party however has the starting point for their route construction but does not know what to do and so the collaboration must start immediately for both to share a common start and a common goal. The diagrams below outline the standard initial map configurations. The first in Figure 4-4 illustrates the visiting places and predetermined routes of one party; the second in Figure 4-5 illustrates the starting point for the other.

Participants have to cooperate to find a mutually satisfactory solution. The task for the participants is to achieve mutual orientation through the map resources and have the same route-plan for all practical purposes at the end. The designed problem is that the maps are potentially ambiguous because they are different versions, and the contingency of the maps as adequate representations emerges locally in the dialogue. See the examples in Figures 4-6 to 4-9. The task is not a naturally occurring one, it has been fabricated to draw upon the everyday methods of human actors - it should be easy to accomplish and require no special training - and to involve the participants in completing what is essentially a visual task through a modality with no visual contact. The task is an abstract one in that a generally available and commonly used representational resource is being used to plan a potential set of actions but the maps are not being used as resources for the physical traversal of a real city. However, given the nature of representations, the maps are essentially inadequate for finding the 'real' thing, ie. what is being represented. The sense in which the map is about a particular city and its spatial and social features must be constructed from the visual and textual surface of the map, and achieved in local and circumstantial agreement within the talk and the CMC dialogue. The following examples in Figures 4-6 and 4-7 show the differences between the two published maps around the starting point shown on map B. The examples in Figures 4-8 and 4-9 show the problem for the party using map B in locating Cherry Hinton Hall and finding a route that parallels that on map A.

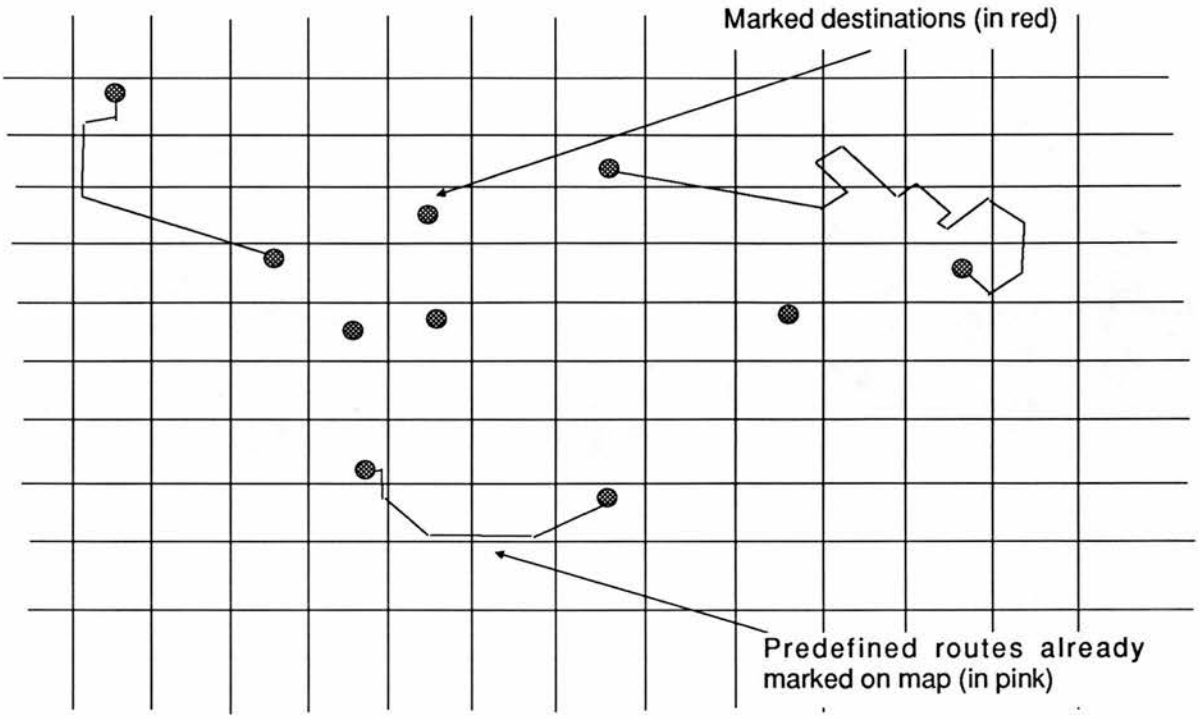


FIGURE 4-4 - Grid scale and markings on map A

FIGURE 4-5 - Grid scale and markings on map B

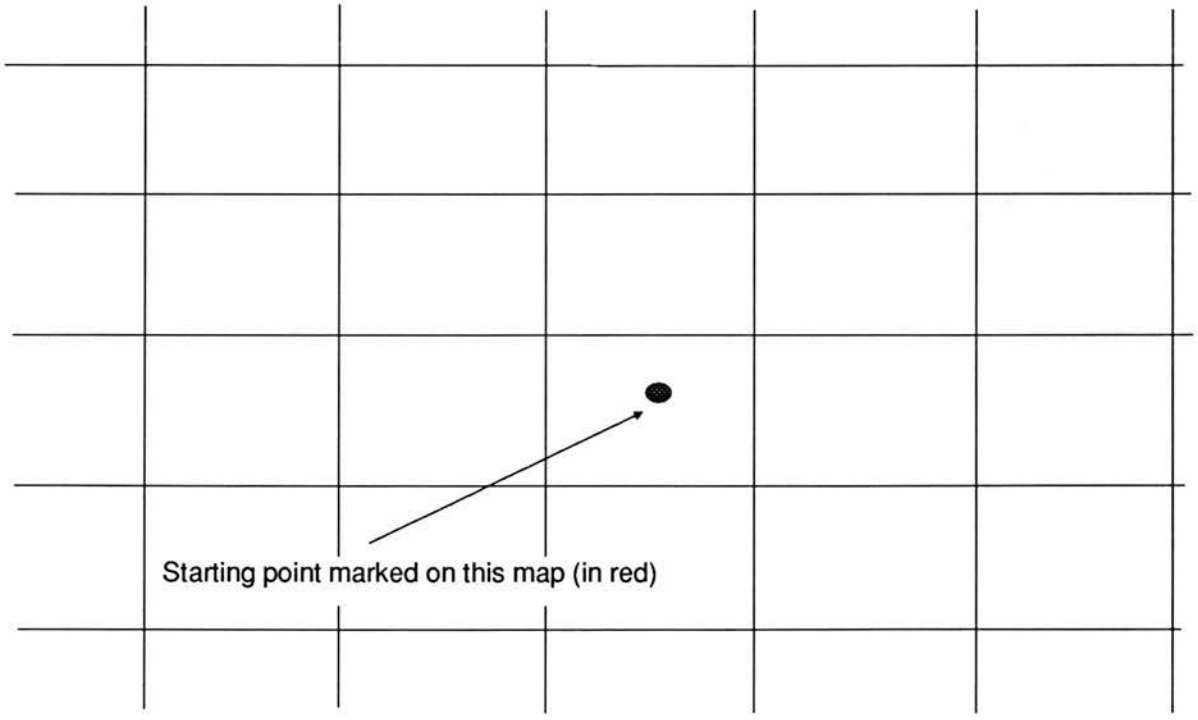
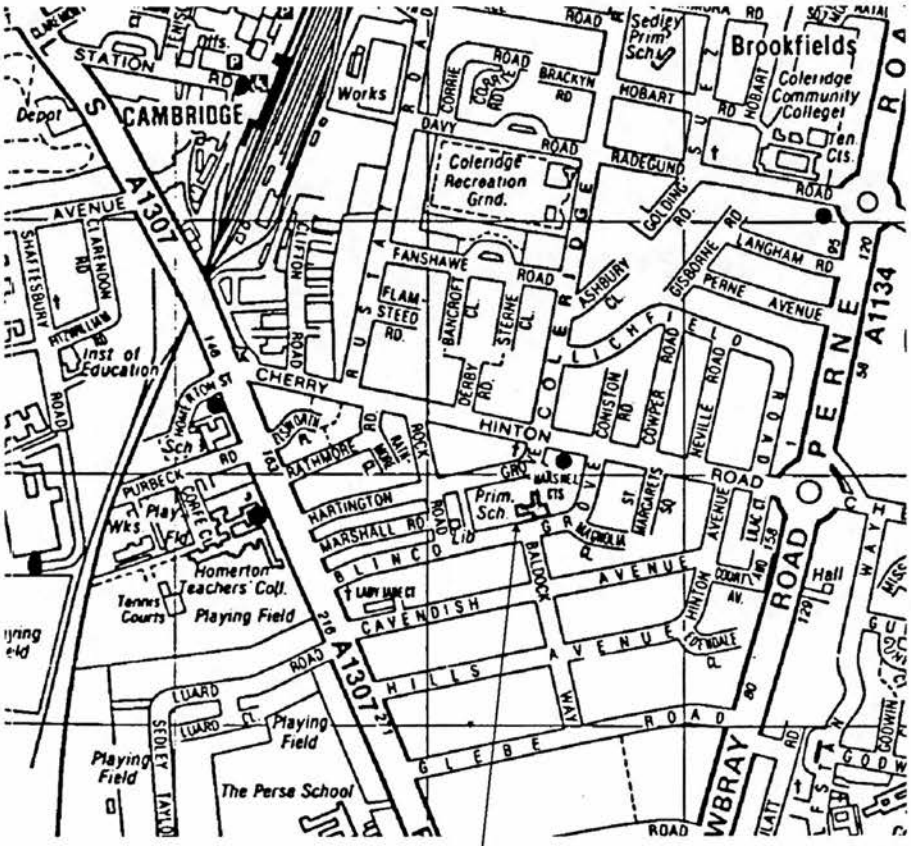


FIGURE 4-6 - A selected portion of map A showing the starting destination.



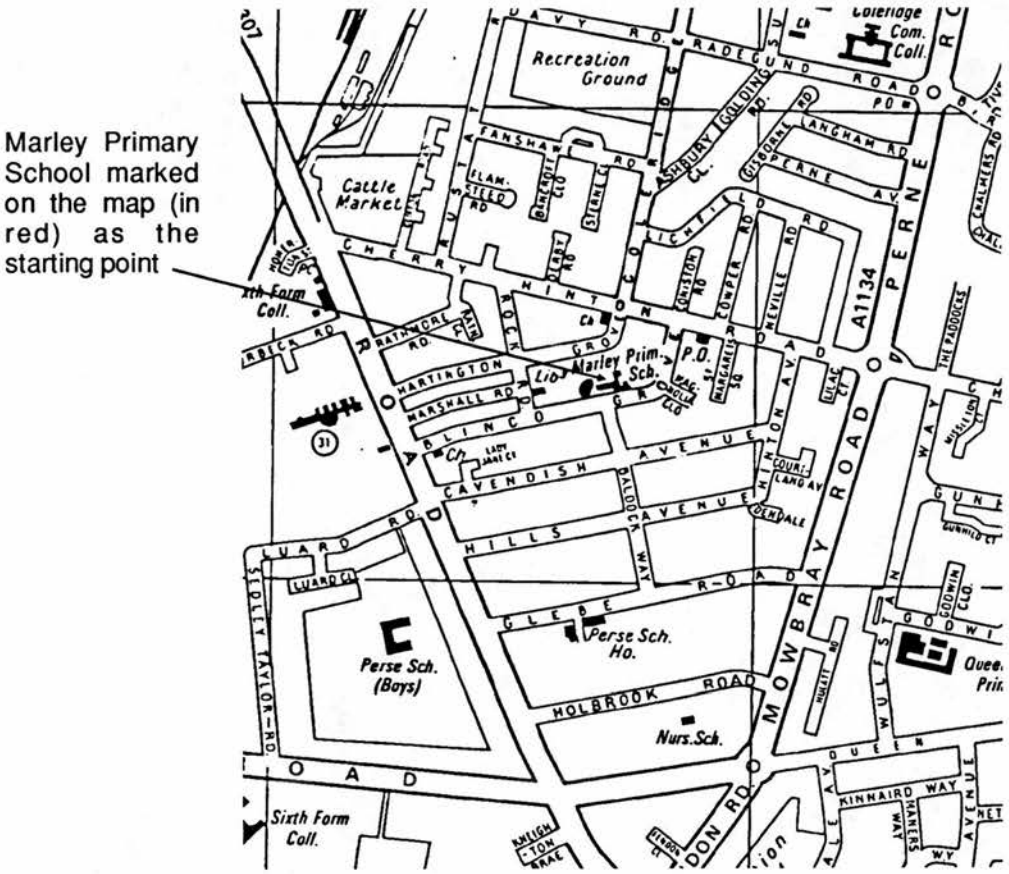
Maps and plans have been chosen because they occur in many practical settings, and draw upon practical reasoning in their interpretation and use. They are produced, whether abstract or occasioned, to be recognisably about a 'shared world in common' through practical reasoning. They are also used because they allow a concrete reference to the materials used in achieving a task.

4.4.3 Subjects

The selection of subjects was not done on a standard rigourous experimental basis, however some care was taken in filtering the subjects. The candidates for taking part where chosen from three sources: graduates in the University, secretaries working within the University, and others outside the University. This was done deliberately in order to locate potentially

interesting types of users of the modality, eg. computer-literate graduates, secretaries skilled in typing, and outsiders who felt they had neither of these skills. The requirements for taking part were that the candidate was a native speaker of British English - the participants were

FIGURE 4-7 - A selected portion of map B showing tthe starting destination.



mostly from Scotland and England. It was felt that a primary corpus of British English speakers would engender conduct closer to the author's own experiences as a life-long member of British society. Thirty three people took part in the experiments, of which three had to be found at short notice because of absences. Subjects were assigned a slot in the schedule. During the experiment, it emerged that some of the participants knew the city represented on the map because they had stayed there for a period of time. This did not affect the study, as there was no control placed over the prior knowledge of the participants.

4.4.4 How the Study Was Run

A form explaining the nature of the study and what would be required if they took part was distributed to each candidate prior to the actual day. The form briefly said that the aim was

FIGURE 4-8 - A selected portion of map
A including one problematic destination.

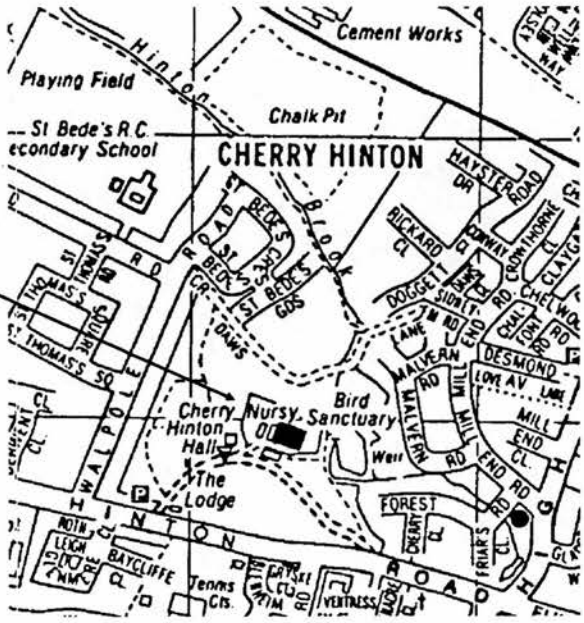
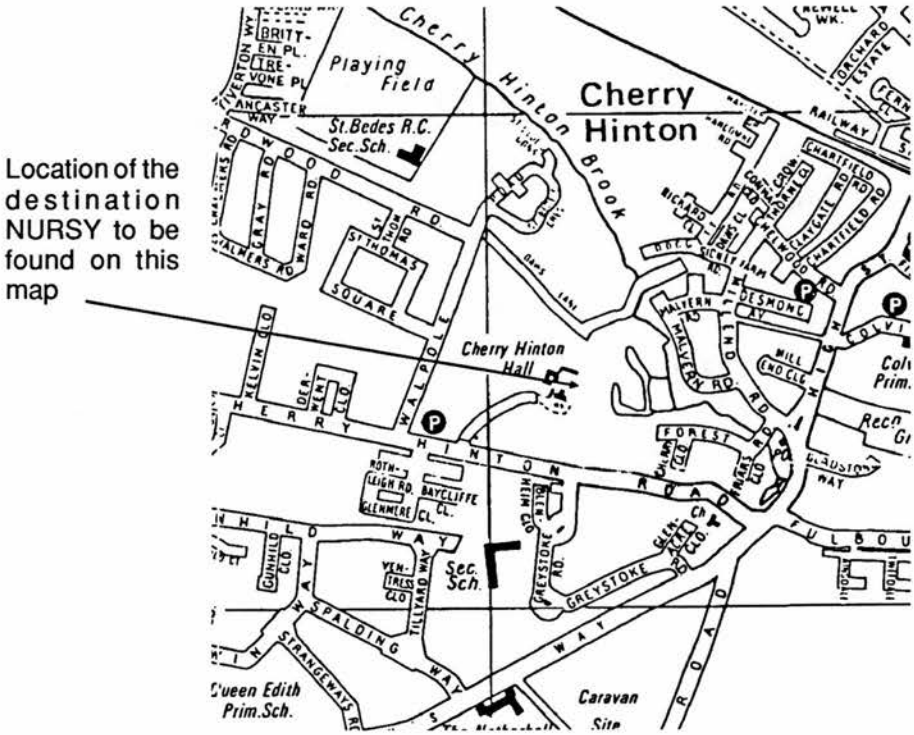


FIGURE 4-9 - A selected portion of map B
including one problematic destination.



to look at how people work together on a task, especially when they had to communicate in different modes. The task would be simple, requiring no special training, and it would not be a test. The possible modes they might be asked to use to complete the task were listed as: keyboard which is like using a computer terminal; face-to-face which is like ordinary conversation; and telephone which is like talking on the phone. They were also told that the experiment would be recorded on video and audio tape so that a record of the events would be available for study, and that confidentiality would be preserved.

Before the experiment the participants were divided into two sets A and B. Just before the experiment was formally begun, the participants were given a sheet of instructions appropriate to their assigned role and mode in the experiment. Irrespective of the mode under study, the party in set A, who were in control of the destinations and fixed routes on map A, were given the following instructions:

> ----------INSTRUCTIONS----------
> You have a map or plan. The other person also has one, though it may be slightly different from yours. Imagine you are walking and that you may have to use the map in the real world. The other person has a start point marked for a route to begin at. The other person has only been told the above, but not the instructions below, and therefore doesn't know what to do. You will have to explain, and work this out together.
> -------
> On your map is a pre-drawn route (in pink) that must be followed, and some destinations (in red) that must be visited at some point. You have to work out with the other person how to follow the predrawn routes and visit the destinations. The aim is for both parties to have a route marked on their maps so that if they had to use the map in the real world then they would both follow the same route. The start point for this route is marked on the other person's map. From the start point you must trace a route (in pink) that visits all the destinations and follows all the marked routes, and then returns to the start point. The other person should do the same and you should both agree as best as possible.
> -------
> A time of half an hour will be allowed. Don't worry if you finish early or don't complete the route in the time available. It is not a race or a test. Just act as naturally as possible !!
> ----------

The other party in set B were in control of the start point, but did not know what to do on map B at this point. Their instructions were as follows:

```
----------INSTRUCTIONS----------
You have a map or plan.  The other person also has one,
though it may be slightly different from yours.  Imagine
you are walking and that you may have to use the map in
the real world.  There is a start point for a route marked
on your map (in red).  If you have to mark a route then
use the pink marker pen.  The other person has been given
extra instructions and will explain what you have to do.
It will require a lot of cooperation. A time of  half an
hour will be allowed.  Don't worry if you finish early or
don't complete the route in the time available.  It is not
a race or a test. Just act as naturally as possible !!
-----------------------------
```

With respect to the mode that both parties were to use to complete the task cooperatively, one of the following instructions was given:

```
----------INSTRUCTIONS----------
1) face-to-face: You are in collaboration with the other
person.  You are responsible for the route on your map,
but you can look at the other person's map and route if
you want.
2) verbal: You cannot see the other person's map, but you
can talk as much as you want.  It is as if you are on the
telephone.
3) keyboard phone: You cannot see or hear the other person
but you can communicate by typing to them through the
terminal keyboard.  Use only the normal typewriter keys on
the keyboard.  The special keys to delete characters and
move down a line will be demonstrated before the
experiment begins.  What you type will appear on one half
of the screen and what the other person types will appear
on the other half.  You can both type simultaneously.  If
there are two of you working together on one map then you
can talk to each other as normal and use the terminal
link.  How you do this is up to you.
-----------------------------
```

In the coparticipant 'constructive interaction' experiments, the seating arrangement was decided at the time by themselves. The seating meant that certain tasks and roles would be necessary for that person. After a reading of the instructions, problems and questions were resolved until they were happy to continue. The experiment was started after the recording equipment was ready, and continued without interruption until the task was completed, or else terminated by the experimenter because the time had run out. The only person present during the experiment was a recording technician in one of the rooms. The equipment and the technician were hidden by a screen from the participants.

4.5 Results

In summary, an empirical investigation of the achievement of dialogue in a synchronous, computer-mediated modality called keyboard PHONE (KP) makes up one half of the empirical data corpus. In the communication modality KP, participants could communicate with each other in different rooms by typing at their terminal as if in a shared visual space. The computer-mediated modality was the only means for the participants to communicate and cooperatively complete the task. Also, a short empirical investigation of co-present collaboration and the local, contingent coordination of concurrency and reciprocity in the flow of conduct was carried out in parallel. This second strand emerged as a relevant direction during the CA style interaction analysis of the first half of the corpus[9], and is reported in Chapter 6.

The next sections will describe how relevant features of the conduct were made available as data, given that the construction of inscriptions by participants is not reconstructable from a script or residue of those inscriptions, and moment by moment copresent activity of coparticipants is not simply available to a present observer. The whole procedure will be grossly simplified as follows. First, an audio/visual recording was carried out, that made some aspects of the conduct available for future inspection using electronic technology, including a trace of computer-mediated events. Then, the recording was examined and transcribed from audio-visual records to the graphic medium using a notation convention. The notation convention is a set of graphic symbols for representing relevant features. The conduct was made visible as a graphic record, ie. a readable representation of some features of the recorded events. Decisions that were made about this procedure will now be examined.

4.5.1 Data Collection

A fundamental principle discussed in Section 4.3 was that some means of capturing the visual and audio features of the conduct is necessary. Audio-visual recording was used so that

9 In addition, small studies of copresent, telephone and keyboard interchange modalities were conducted but these are not analysed in this thesis. For copresence, the parties were in face-to-face contact and could look at each other's maps. For the speech telephone, the parties could only speak to each other. For the keyboard interchange, the parties were linked by a computer-mediated restricted 'flip-flop' modality that allowed only a strict alternation of turns, in which turn allocation was under the current speaker's control.

relevant 'visible' aspects of the conduct could be viewed and reviewed as many times as required. Additionally, the event is available for inspection by others, and so the lived-work of analysis can be repeated by others. Two main things to record were: the communication between the parties in the CMC modality; and the activities involving the chronic process of monitored production and interpretation by the coparticipants. The set-up as described in Section 4.4.1 resulted in two separate video records of different aspects of the cooperative activity being recorded. Later, the video sound was transferred to audio tapes to aid close repeated listening. The following key will help in identifying a particular example from the corpus[10].

KEY for referencing experimental events:

1+1	Number of coparticipants,
?-?	Number of participants in each party across modality,
KP	Virtual dialogue space, keyboard PHONE,
C	Copresent collaborating coparticipants.

The reference code found in the data fragments is based on the codes M1 - M10 that identify the separate experiments making up the corpus.

4.5.2 Transcription and Record Development

A video or audio recording is traditionally rendered into an interpreted symbolic form using a graphic notation system. This process should not replace the original transient recording that must be watched or heard, but by rendering the conduct strange and unfamiliar it does help us to reveal the interpretative practices, etc. that we take for granted in finding the sense of everyday conduct that we see and hear. Procedures for doing this have evolved during my experiences with transcribing in the interests of analysis, and are largely drawn from Gail Jefferson's CA conventions, as described in Jefferson (1983), which are adopted by many other practitioners. Some aspects are different, particularly the occasional representation of prosody and stress, which is borrowed from linguistic work on prosody. The record style used for graphically presenting the primary transcription, the record from which others are derived, is complex and will be described below. The linear turn-based record style of CA is also used but it will only be illustrated towards the end of the chapter. An explanation of the

10 The statistics are: for (1+1)C, 11 units or 91 mins were transcribed out of 300 mins; for (2-2)KP four units of 39 mins were transcribed out of 110 mins; for (2-1)KP three units of 13 mins were transcribed out of 80 mins; for (1-1)KP one unit of 25 mins was transcribed out of 25 mins. The statistics for the amount of data in the corpus is stated in terms of amount of time and units, eg. a unit is one complete session. Note that transcribing (2-2)KP involves ten streams of synchronised activity and takes approximately 'times 100' to transcribe properly.

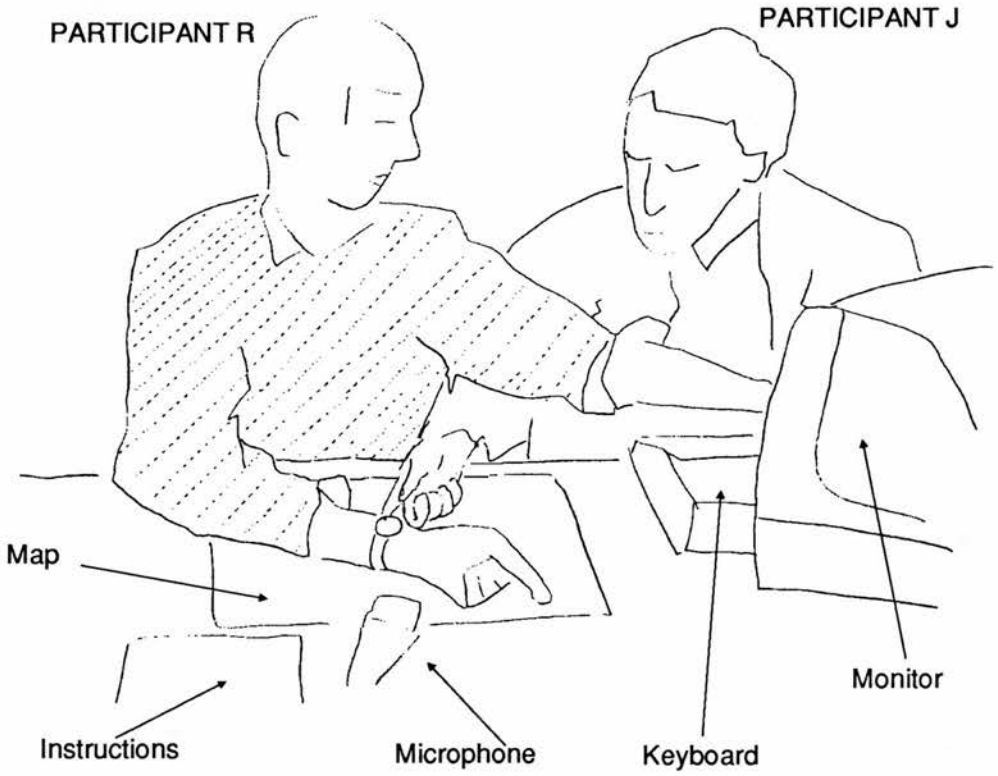
notation conventions now follows in three sections: audio, keyboard, and the body. Appendix A gives a list of all the notation conventions and a brief explanation.

First, the audio stream is transcribed in the 's' stream on the graphic record, which includes speech and other relevant noises in the local environment. Speech sound is represented in standard orthography, though some pseudo-phonetic representation is given if relevant. On the turn-based records speech conduct is in the **HELVETICA FONT**. Intonation has been recorded if clear, but has not been if the notation does not allow adequate description of the speech. In this case verbal comments come after the speech fragment indicating quality. Four types of contour are recognised: falling intonation (.), rising (?), continuing or indeterminate (,) and exclamatory (!). Unfortunately, the development of an accurate prosodic notation system suitable for interactional analysis has not proceeded very far. Also the following features are transcribed: extensions of sound (:), cutoffs (-), loudness (CAPS) and softness (°), mouth clicks (.t) and audible breath characteristics (.h, etc). A quotation or reading voice is noted ({...}) as this is common in the coparticipants talk about the dialogue in the CMC. Temporal measures of an absence of activity are given to the nearest 0.5 second except for micropauses '(.)' of less than 0.2 seconds. In other cases, they are simply marked as short or longer pauses of no specific length, eg. (pause) or (longerpause). Additionally, the spacing between items in the stream relative to the pacing of other activity indicates the approximate length of the inactivity. Difficulties in speech transcription, ie. doubts over the identity of the speaker, or over the recognition of speech or sound are marked in parentheses. If a speaker stops and immediately a distinctively different utterance or speaker continues, then this is marked using '=' at the disjuncture on the turn-based records. Onset of overlapping speech is traditionally marked using '//' or '[' and this occurs when data is presented in the traditional turn-based way. However, in the records developed specifically for this study overlap is intrinsic in the graphic representation of streams of simultaneous conduct. Records are to be read with time proceeding horizontally with multiple streams of conduct sandwiched so that vertical alignment with other streams on the graphic record means simultaneity in time. Thus for the coparticipants any speech conduct by them rendered in the two horizontal streams that is in vertical alignment on the graphic record is in overlap, ie. they are saying something at the same time. This will be illustrated when a reading of a fragment is demonstrated.

Second, activity in and through the physical keyboard and screen in the achievement of dialogue in the CMC must be transcribed. Figure 4-3 in Section 4.4.1 illustrates the screen display. All operations on the keyboard that produce characters or relevant effects on the display are recorded in the 'k' stream. On the turn-based records the keyboard activity is distinguished from speech and bodily activity by using the COURIER FONT. Key presses

produce characters on both displays almost simultaneously so recording of the screen activity needs only an indication of author and not a distinction between time of transmission and arrival. Characters appearing on the screen are written down in their symbolic form, as well as attempts to generate a character ([ ]), to erase characters (within '//' and deleted in reverse

FIGURE 4-10 - A sketch of the video image from [M5:1'48"].



order to typing) by pressing the delete key, and to move the cursor out of the implicit linear horizontal progression of next places to type and around the display. In addition, the space bar effects the spacing of letters into perceivable word groups (•) and the return key creates new work space (®).

Third, non-verbal or bodily activity is an integral part of the embodied collaborative activity. The sketch in Figure 4-10 illustrates that gesture and gaze are potentially relevant in this collaboration and for locating the organisation of activities that gives meaning to the inscriptions in the CMC. This sort of activity is represented in the record in the 'g' stream. Any appearance of a symbol read from left to right implies a new observable phenomena, or a change from an action or activity currently operative.

Eye gaze is distributed in general between the map (m) and the screen (s) for both participants. Mutual gaze (e) occurs extremely infrequently, though this does not imply

inattention. Gaze that is not to the usual domains of activity is either explicitly indicated or marked as neutral space (Ø). Attention to gaze is made available through peripheral vision and bodily orientations are continually displayed in the video recordings. The coparticipant who types also pays much attention to the space of the keyboard (k). (See Sudnow (1978) for a fascinating ethnomethodological investigation into learning jazz piano playing focusing on the hands and the keyboard.) Because gaze generally continues at the same place for a period of time it is only marked with a symbol once, at its acme. That is, when a person's gaze moves to an object or in a particular direction the moment it reaches a stable position it is notated; it can be assumed that after that moment it remains there until the occurrence of another gaze symbol. A gaze movement that does not stabilise or is interrupted is marked in parentheses.

Gesture is more complex. Participants commonly point at the map and the screen and gesture in particular ways towards either location or in neutral space. The onset of a gesture is marked by '((' and the closure, if it can be determined, by '))'. Standard symbols have been developed to represent the acme - point of maximum energy - of a general form of gesture to be found in the corpus, eg. pointing at the map or screen, gesturing to the map, screen or neutral space. Descriptions of the place or movement of the gesture can be described in square parentheses ([...]). Because of the possibility of left/right hand simultaneity and sequencing, prefixing an 'L' or 'R' to the symbol indicates the hand doing the gesture. A single vertical arrow (↓) indicates a sharp point or jab to the map, and a double barred horizontal arrow (⇒) indicates inscribing activity on the map surface. Incomplete gestures are enclosed in parentheses, and small movements are within '°'.

The graphic records themselves, on which the notation symbols are juxtaposed, were specially designed to graphically render details relevant for the study. The transcription records are organised to represent conduct as a linear temporal stream that traces simultaneous actions, placed on the record in relation to the keyboard activity that is the party's only mutual reference point. Both the parties' dialogue and coparticipant speech and non-vocal actions are recorded using linear channels that are graphic records of the transcriptions of particular types of events, eg. speech, gaze, gesture, keyboard/terminal actions. Important temporally simultaneous events are recorded through vertical cross-channel alignment. There are some differences between the linear multilayered representation of the temporal stream of conduct and the traditional records adopted by most work in CA and discourse analysis. In CA the turn-based records are oriented around the notion of the 'floor' - coordination of speaking - so as to display the temporal evolution of the focused activity of speaking that is being represented. Some research in CA is keen to represent simultaneity, and non-speech actions are annotated in parallel with a

speaker-segmented, linear speech record. In DA a standard orthographic 'interchange' record of the speech stream is used in which overlap is written out because of an interest in the 'words'. Because of the complexity of the records some illustrations of how to read them will be given.

In Example B1, which can be found in Appendix B, only two participants 'face' each other across the modality, A and J. Normally only the wider band (k) that records the CMC events is notated, but in this case, and unintentionally, the gaze of one of the participants was recorded on video and is also transcribed in band g. The temporal flow of the activity runs from left to right. Notations of important time intervals, like pauses, are marked between vertical bars. The notation conventions are in Appendix A. This example begins at the start of the experiment so that neither have had any communication through the modality yet. After a short period A begins to type the message "can you tell me the starting point ?" and presses return, thus a new line is begun. During this activity, A deletes two characters, "he", and gaze is directed to both the keyboard and the screen. After pressing return A gazes at the screen. After a short while J responds with "i am at Marley Primary school". A continues to gaze at the screen for 6 seconds after the last visible change to the window by J, except for a brief glance to the map during the typing of "school". The dialogue continues but we will look at different example.

Example B2, which can be found in Appendix B, represents the relevant features of two pairs of co-participants collaborating on their textual map surface, and communicating with the other pair in and through the CMC modality. Unlike most CA transcription records there are 5 horizontal bars in the (2-2)KP record, and they are collected together in a bundle that represents the following simultaneous features. The middle bar 'k' represents the shared communicative space mediated through the keyboard. All other activities by the coparticipants are mapped out in relation to this shared stream - one coparticipant pair above, the other below. Above and below the middle bar are two pairs of bands that notate the speech 's' in one, and bodily activity 'g' in the other of each co-participant pair. All five bars represent the flow of time from right to left in a visual way. A slice through the five bars represents a moment in time, by which synchronies and concurrences can be mapped across modalities and participants.

A reading of this example will follow. At time 2'16" in the record, K is looking at the map surface and G is looking at the keyboard as K's pointing gesture on the map surface reaches onset. K then says: "a region=i mean this has got regions on", and towards the end of this utterance K produces several rhythmic jabs on the map surface. Also, G looks to the screen, then the map, and after K's utterance G replies with "right" and looks to the keyboard. Next she explicates a possible candidate for the next inscription by saying, "so i'll put region,", and

immediately starts typing the construction "Reg" so as to communicate with the other pair. At time 2'16", the other party, E and R, are looking for signs of activity on the screen. E looks to the keyboard and attempts to type but nothing visible shows on the screen. E indicates trouble with the utterance, "o- oh-", but then resolves the problem and finds the appropriate keys to type "Cavendish ". At this moment both parties communicating through the CMC are constructing contributions in overlap, but at this point neither party displays any recognition of this. It only becomes significant to G when she glances to the screen and then says "hang on we got some more what's this?" attracting K's gaze to the screen. The other party carries on: participant E initiates an elicitation sequence with the simple tagged utterance "avenue was it?" and then starts to type "Aven" as R confirms with "yeah".

The reading of a (2-1)KP example is practically the same as for (2-2)KP but simpler. One side of the modality has only one participant so the record structure has only one stream of keyboard activity and one stream of gaze for that person, but the other streams are the same as for (2-2)KP. In the next example, which can be seen in Figure 4-11, the more familiar turn-based record is used. This style appears frequently in Chapter 5 and is used when possible to simplify the presentation of examples. However, the simplification is only performed when it does not misrepresent the conduct in ways that are pointed out in the next chapter, eg. by misleadingly presenting items in the record as if they were simultaneously perceived by all parties. The record interleaves both the keyboard activity of both parties and the speech activity of coparticipants, as well as bodily activity. It is to be read from left to

FIGURE 4-11 - [M5: 20'04"-20'22"]

```
     D:     continue along the footpath along to ®
→            CHERRY HINTON ROAD
     D&M:   ((Both look to screen))
            ((M reads a new contribution by other party, "STOP !!!!!!!!!!"))
     M:     {stop} (.) ur::.
             (pause)
→    D:     ®®OK
```

right and from top to bottom on the assumption that it renders conduct from the perspective of one party or participant as a linear stream occurring through time in the same order as on the page. Thus D types the contribution, "continue along the footpath along to " and types RETURN thus starting a new line on the screen. Then D continues with, "CHERRY HINTON ROAD", whereupon both look to the screen. M reads what the other party has typed, which has not been recognised by either D or M so far, and says "stop" in a characteristic quotation voice. Note that the activities of the other party during this fragment are not documented in

the record. He pauses, says "ur::", and after a pause, types two RETURNS and "OK". Overlaps between different activities are indicated when relevant in the same way as for talk.

4.6 Summary

This chapter has described the details of the investigative audio-visual study of a computer-mediated dialogue modality. After the brief overview in the introduction a survey of related research was given. The study was contrasted with other research on modalities, computer-mediated communication, constructive interaction techniques, task dialogues and map direction-giving. In particular it was found that many studies lack attention to the empirical details of dialogue conduct and use coding and statistical techniques for the analysis. For example, only products such as scripts or codings of the activity in the modalities studied were used in the analysis. This is in marked contrast to the principles of the study here that are borrowed from the naturalistic investigation of conversation. An experimental setting is used only because of the nature of the CMC being studied. The constructive interaction technique generates accounts of the activities as they are engaged in: both the constitutive practices dialogue in the CMC and the collaborative work of the coprarticipants. Another principle is that audio-visual recordings and transcription techniques must be used in order to remain faithful to the participant's perspective. The procedures of the study and the transcription techniques, the first stage in the analysis of the results, were then described in depth. Audio-visual recordings were essential in order to capture the fleeting simultaneity of the party's dialogue and talk, as was a complex graphic rendering of the vocal, gaze, gesture and CMC activity.

In Chapter 5, the full analysis will deal with the interactive achievement of dialogue in the CMC. Participants only have a restricted range of resources for achieving sense compared to everyday copresent talk, and yet they have an open floor of possible mutual participation in which each has the physical opportunity to construct action at any time. The analysis will document details of how monitoring and participation are locally achieved that give the inscriptions in the CMC their sense as dialogue. These demonstrations will serve both to highlight the problems and limitations of current computer models of dialogue and point a way forward for understanding the resources and environments required for richer human-computer interaction. Chapter 6 analyses a complex fragment of the collaborative work of the participants in achieving the task, paying particular attention to the possibilities for sophisticated interactive dialogue artifacts.

Chapter 5

THE SITUATED INTERACTIVE ACHIEVEMENT OF COMMUNICATION IN A VIRTUAL DIALOGUE SPACE (VDS).

And what is the use of a book, thought Alice, without pictures or conversations? (Lewis Carroll, Alice's Adventures in Wonderland)

The pieces are readable, take on a sense, only when assembled; in isolation, a puzzle piece means nothing - just an impossible question, an opaque challenge. (Perec 1987/70, p. xv)

5.1 Introduction

From the arguments in earlier chapters and the empirical methodology described in the last chapter detailed analyses of instances of dialogue 'through' machines will be presented in this chapter. The main aim is to show that interaction is a constitutive domain of dialogue and cannot be taken for granted or ignored in computer models. In particular the analyses demonstrate that communication is interactively constructed in and through a virtual dialogue space (VDS), which is a simulated graphic space in which participants can coordinate their activities as dialogue, that is virtual because the computer supports the image of a visual workspace mutually accessible to both parties who are spatially-distanced.

There are three main activities of interest in the study. First, there is the cooperative work of achieving an adequate mutual route construction. Second, the situated, chronic process of production of the inscriptions in the VDS which is distanced from the other coparticipant. The inscriptions are worked on and worked over to produce abstract contributions intended for an audience. Third, there is the interactive engagement in dialogue by the parties. Normally in writing the second activity assumes primary significance when engaging in dialogic correspondence. Texts are produced which have a durable character and become separated from the saturated contexts of production and the projects of those who created them. In talk

the third is primary; the second does not exist as participants are engaged in contexts of copresence. If inscriptions are produced, then they are joint situated products in a way that texts written by absent others are not. Their sense is bound and embedded in the activities from which they arose. The constituting activities that give inscriptions their sense are also to be found in Grosz & Sidner's (1986) dialogue example set-up but they are unfortunately disregarded in their analyses. For instance, the mediator between the expert and apprentice is doing interesting and relevant work to locate a contribution from the apprentice's speech and type it to the expert, and vice versa. In the VDS, all three activities are interwoven. However, a tension is present between the demands of the second and third activities. This is a divergence from the copresent situation in that many routine resources for mutual activity are unavailable. This results in particular problems for the coordination and interpretation of their dialogic activities. Parties are contingently resolving the relevance of contributions, eg. sequential misplacements occur, as they engage in multiple and simultaneous activities. Visible traces are available as records of the conduct, however, the record of the participant's dialogue is not embedded in copresent activity, eg. a jointly constructed diagram on a piece of paper in front of two people talking. It emerges within the context of two activities having competing claims on attention.

Before the main analyses an extended discussion considers what methods are to be used to recover the situated work that gives the inscriptions in the VDS their sense as constributions to a developing dialogue. Unfortunately the screen residue is only an abstract sediment of some of the events that occurred and turn-based renderings are misrepresentative, so complex renderings of audio-visual recordings of constructive interaction are motivated. The analyses investigate how parties construct their actions through the technology available, how they visually and temporally monitor their own and other's production, how participation is locally organised, how contributions are relevant to each other, and how new orders of participation emerge in dialogue. The main finding from the analyses is that the participants in the VDS do use similar resources to those found in conversational talk, such as sequence or repetition, but fragmentation of the basic features into emergent organisations is routine. For example, participants do construct their participation in similar ways to the turn-taking in conversation though the particular constraints of the modality result in a diffusion of the notion of 'turn' and the emergence of parallel dialogue activities that cannot be accounted for by the rule 'one speaker at a time'. The structures found in the VDS are unusual yet motivated by the situated, mutual work of constituting intelligibility out of the materials and resources that are appropriate at the time. In this chapter the embedding of inscriptions in situated practices essential for their interpretation will be illustrated as part of the situated embodied achievement of dialogue in the VDS. The importance of the context of interaction as constraint and resource, eg. the permanance-transience properties of the medium, cannot

be overemphasized. In conclusion to this chapter, the relevance of the study to computer models will be discussed, which will lead into the next chapter which considers interactivity and dialogue artifacts.

5.2 Recovering Practice

This section will illustrate a number of issues that will be developed in this chapter. It will suggest how to investigate and recover the practice that empirically weak interchange models of dialogue such as Grosz & Sidner (1986) have taken for granted and missed. The recovery will be accompanied by the development of an alternative method of rendering details of conduct that gives access to the situated work of achieving intelligible dialogue out of the materials of the VDS itself. A motivation for the 'constructive interaction' technique is given. Lastly, some phenomena of the virtual dialogue space will be illustrated that are analysed in depth in later sections.

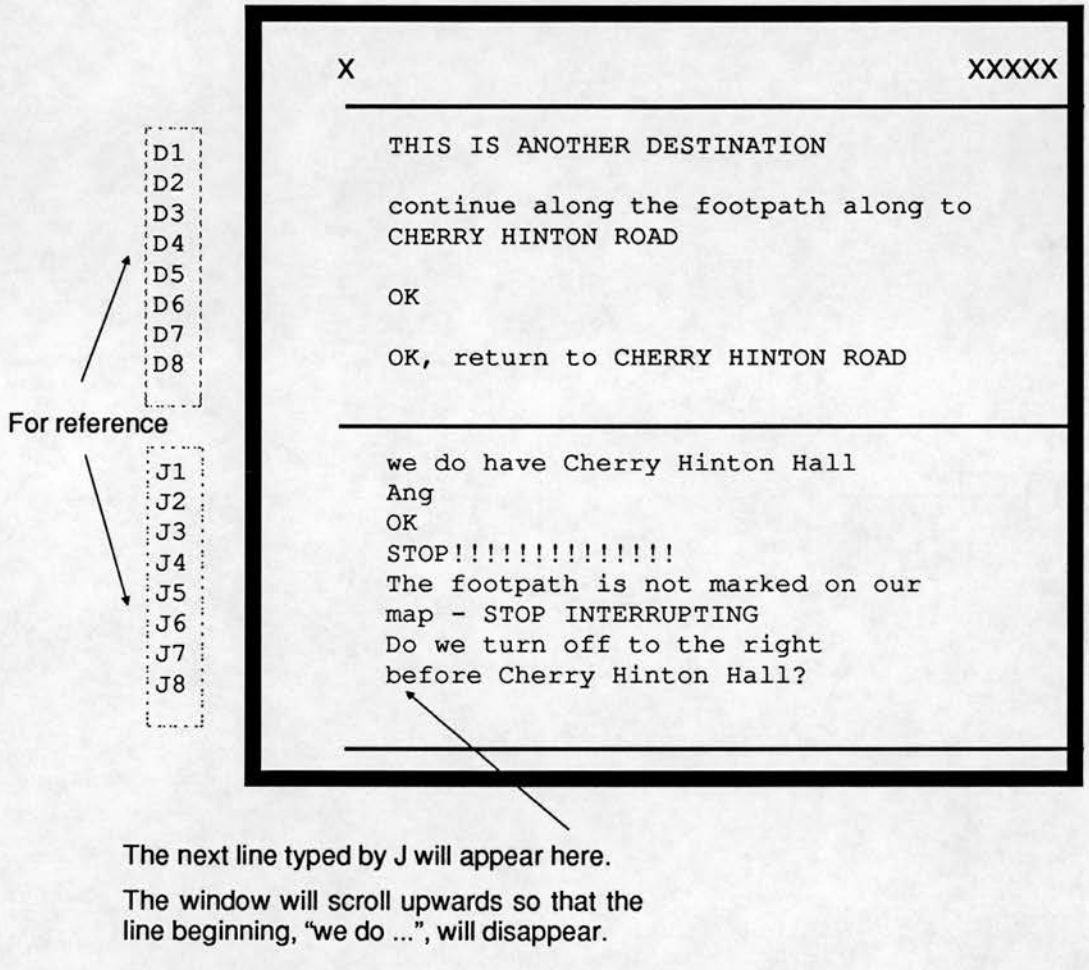
5.2.1 Towards an Adequate Rendering

In order to reveal the empirical problems with Grosz & Sidner's account it will be rewarding to take a close look at how an abstract conception of dialogue misses the situated details that are demonstrably real for the participants in the dialogue itself. The crucial questions for any account are just what is 'the real stuff' of dialogue and how it can be represented in some way so as to be talked about. Unfortunately, the situated production of dialogue is usually ignored in favour of a textual or dramatic 'interchange' model that supposedly represents the 'real substantive content'. In many computer theories of dialogue, including Grosz & Sidner (1986) and Reichman (1985), this model is implicit in how the claims of the theory are illustrated.

Let us examine some examples from the study that will clarify what is lost in the process of rendering data according to the interchange model. First, an example of what the participants in the dialogue actually see while engaged in dialogue is given in Figure 5-1. This example shows the visual appearance of a sedimented screen display at a particular instant in the course of the dialogue activity, ie. over a period of time the contributions that were constructed on the fly in the VDS come to be layered on the display much like the geological formation of sedimented rock strata. Both parties could see this visual display screen at the same time and in the same form. It is the textual surface on which coparticipants construct and interpret contributions. It can be treated by the parties as a living surface of ongoing work, or it can be just a record or residue of events that occurred during the course of engaging in dialogue through the modality. However, the screen display is not an accurate document for the purposes of analysis because the significance of items or objects in the display is only to be

found in the dialogic practices that engendered them, and this is why audio-visual methods

FIGURE 5-1 - The sedimented screen display of M5 at 21'26"



of analysis were used here. There is a simple temporal history recoverable in this example, ie. events higher up in the windows occurred before those below them, but it is not quite true that the display contains a sediment of all the events that have taken place, as there are some events missing from the record. Reconstruction by the analyst of what actually happened, eg. the sequential organisation of 'turns', could be attempted from the screen residue by drawing on the findings of previous analyses and commonsense. A story could be constructed, much in the vein of Garfinkel's 'et cetera' clause experiments that demonstrated the indefiniteness of accounting procedures (Garfinkel 1967). But problems will occur, for example, a repair initiator by D in response to line J2 is not recorded at this later time because it was almost immediately deleted. Also there are lost lines, eg. sequentially relevant lines to those of the other party are missing from the display before D1 because of D's policy of double spacing 'turns' or responses. It is interesting to note that the coparticipants face a

similar, though practical, problem when they sometimes refer back to the display to recall or try and reconstruct a past event that is relevant once again. However, the display is more meaningful to the participants in situ because they have been a part of the process of constructing meaning of which the display is but a residue. This illustrates the much forgotten fact that remembering is not completely a mental process, but is distributed across people, artifacts, and the world. Recall and memory is as much 'out there' as 'in the head', eg. in this case, the configuration of the display is essential to the participant's remembering of what happened. Also, aspects of the production of contributions, ie. their performance, that may have been relevant to the evolution of the dialogue are not implicitly represented in the screen documents themselves. For example, the pace of typing and hesitations during the construction of inscriptions as contributions to the dialogue are orientable performance characteristics that are not recoverable from the record. Recipients can shift orientation to different activities if a hesitation occurs in the construction of the current contribution. Also, the speed of typing can vary from slow to fast, and there is the keyboard repeat mode in which a character is repeated very fast on a single line. In one instance, an exclamation character is reproduced portraying a sense of urgency in order to stop the other party interrupting a construction.

In order to remedy the problems for the analyst who might rely on the emergent record of inscriptions carried out in the course of dialogue activity, a simple rendering in traditional turn-based script form may appear at first to be the answer. Example 1 illustrates, in a particular way, the dialogue sequence leading to the display state in the prior example, eg. [M5: at 21'26"]. D is typing for the pair D and M, and J is typing for the both J and A. Basically, this example is an abstract serial record, ie. the lines on the page can be read as turns in discrete chronological time and they are labelled linearly down the page. The bracketed codes refer to the original place of the contribution on the screen display in Figure 5-1. In addition there is some attention to chronological temporal simultaneity, ie. recording of the clock time when an 'overlap' of two responses occurs. Unfortunately, this form of transcription misrepresents the participants' 'reality' in the VDS because they are not in co-presence as in conversation. Mutual monitoring by 'speakers' and 'hearers' is not a routine accomplishment in this case, ie. contributions are not mutually monitored because of a lack of resources and thus the transmission and reception of entities in the VDS becomes relevant. This means that the relevance or significance of contributions for the participants is not recoverable from this new record because alternative interpretation strips cannot be represented in this rendering. For example, some turns are never noticed, eg. on line 5 the echo repair initiating response "ANG?" is never noticed by the other party; some are not noticed until later, eg. on line 2 the mistyped construction "Ang" is not 'received' until after responding in lines 3 and 4 to the prior contribution in line 1. Also, lines 10 and 15 are not

noticed until after the end of the recipient's current turn construction. Turns are quickly deleted and removed from the record: line 5 does not exist in the record because it was typed and almost immediately erased by D. It becomes an incidental comment for those attending to its production. Contributions are edited in progress, eg. in line 3, "east" replaces "right", line 5 was erased by D before typing line 7, and note that lines 3 and 4 do not appear on the screen display because they were scrolled off the display by D's double spacing of responses. Thus an abstract serial record of the chronological events misses the situated work of constructing contributions and misrepresents the relevances of contributions to the participants in the evolving dialogue.

Because conversation analysis (CA) has dealt with focused encounters between two or more participants, the rendering methods have developed to cope with a temporally singular locus of mutual activity. Participants are actively engaged in achieving conversation as accountably having one speaker at a time and one floor of speaking. Thus a transcript can

EXAMPLE 1 - [M5: 19'04"-21'26"]

A simple 'turn-based' transcription that is misleading.

```
1   (J1)      J:    we do have Cherry Hinton Hall
2   (J2)      J:    Ang
3             D:     [The nusery is to the east
4                   of cherry hinton hall.
5             D:    ANG ?
6   (J3)      J:    OK
7   (D1)      D:    THIS IS ANOTHER DESTINATION
8   (D3)            continue along the footpath
9   (D4)            along to CHERRY HINTON ROAD
10  (J4)      J:    [STOP           [!!!!!!!!!!!!!!!!
11  (D6)      D:    OK
12  (J5)      J:    The footpath is not marked on our
13  (J6)            map -
14  (D8)      D:     [OK, return to CHERRY HINTON ROAD]
15  (J6)      J:                      [STOP INTERRUPTING
16  (J7)      J:    Do we turn off to the right
17  (J8)            before Cherry Hinton Hall?
```

be read as a stream of speech conduct mutually attended to by parties ratified as speakers and listeners. Gesture and gaze are notated around this singular locus and in conversation it is argued that interactants are

> *viewed as simultaneously engaged in fine-grained real-time coordination of speaking turns tracked predominately in terms of surface structural features.* (Heritage 1989, p. 26)

But, the transcript example above cannot be read in this way. In the virtual dialogue space, mutuality begins to break down and the notions of speaker and hearer are not so relevant,

ie. the parties in the VDS do not attend to each other alternately as speakers holding a floor of mutual attention. How and why this occurs is dealt with empirically in later sections. This poses a problem for computer models of dialogue in that the integrity of the concepts of speaker and hearer are assumed in the dialogue script examples used to illustrate the model. For Grosz & Sidner (1986) it is even more of a problem because the notions of speaker and hearer cannot be applied to the apparent contributors to a mediated dialogue in which the expert does not hear the apprentice nor the apprentice speak to the expert, ie. the apprentice speaks and hears while the expert types and reads, both mediated by a human translating between the mediums.

So, a rendering that allows alternative interpretative strips to be recovered is needed, and this can be done by notating more than just a 'turn-based' chronology of screen events. Continuing with the same data fragment as before would have illustrated the argument for the same case, but this is not done because of the length of the final record that would be required to cover the whole fragment [M5: 19'04"-21'26"]. Instead, aspects relevant for how they are addressed by the participants themselves are notated in Example B1 given in Appendix B. However, this example is too simple for the purposes of this analysis. Participants do demonstrate orientations to each other in their contributions - in fact, that is the only way they can display understandings - and thus to the analyst, but because of the restricted modality a number of questions about the situated practice of constructing and interpreting contributions is unavailable to the analyst if one to one participation is studied. What happens when overlaps occur? Why do particular observable dynamics or behaviours occur? Also, how do the participants interpret the other's conduct? What chronic processes of monitored production are involved in constructing a response? These are questions that have no easy answers if normal observation is relied on because the resources for the analyst found in observation of the conduct are much restricted by the modality. Of course, conceptual theorising could take over at this point, but a different and more acceptable way out is adopted in Example B2 in Appendix B.

'Constructive interaction'[1] involves the use of coparticipants engaged in the collaborative work of what one participant did before. Audio-visual techniques are required that are described in Chapter 4. A special transcription method is used that critically renders the embodied work of the coparticipants in addition to the temporal development and simultaneity of activities. From a record of these things the analyst can isolate the situated practices that

1 The technique called 'constructive interaction', described earlier and motivated here and in Chapter 4, is implicitly drawn upon in the analyses of the prior examples. Some of the analyses could not have been made without this method.

give sense to the inscriptions as contributions produced and interpreted in the peculiar circumstances of an evolving dialogue. An advantage can be gained by using a restricted modality. The modality allows the use of the constructive interaction technique because two people can collaborate on the dialogue without too much distortion to the VDS itself. This is not naturally available in normal conversation. For example, imagine constructing an experiment with a pair of subjects acting conjointly and trying to do the job of one person in a conversation, so that a display is engendered of the collaborative and observable work done by the pair in achieving conversation with another person. This is not a suitable strategy because of the unavoidable disturbance to the natural character of the conversation, eg. the mutuality of conversation would be radically altered while the coparticipants worked out what to say to the other person. But, for restricted modalities involving spatial or temporal distance and for ordinarily solitary activities it is a useful strategy for effecting talk of and about the activity, ie. the coparticipants talk about and thus display the sense of what they are doing.

In the example, the chronic process of monitored construction of a response in the modality is now available as a resource for the analyst; but not for the other party. Thus we can explain why the fragment "Reg" appeared in the dialogue space because the collaborative work done to construct the response is demonstrated in the talk of the coparticipants. Owing to the visibility of contribution construction, G noticed the other's simultaneous response and attended to it, thus leaving a half-constructed 'overlapped' response, viz. "Reg". Through G's talk, body and gaze in coordination with displaying accountable actions to the coparticipant K, G's orientation to three sets of activities is visible. The first is the mutual work of constructing the sense of a next contribution. Then the shift to the second, with "right" and a gaze movement to the keyboard, which is to realise that sense as a contribution in the VDS through the keyboard. The potential development of the inscription "Reg" is explicated by their mutual work and G's utterance "so i'll put region". Finally, the third is attention to the dialogue itself, with "hang on" and gaze directed to the screen, as a shared space of understandings. This example illustrates the tension between the chronic process of the production of 'messages' or contributions and the mutual engagement in dialogue itself. Participants G and K are collaboratively engaged in finding a reasonable response. G is also actively engaged in constructing their next contribution to the developing dialogue. But G notices a potential contribution by the others, and thus their own contribution becomes suspended by the relevance of attending to the sense of the "more" that has now been noticed.

The temporal development or performance of a contribution is publicly available as a trace in the VDS, but the coparticipants' talk and activities of looking, searching and planning are not available for inspection by the other party. So, there is a drastic reduction in the mutual

access of participants to each other's circumstances and activities. In addition, the performance of contributions can be treated as a transient event with an observable course, or as a permanent record or inscription. Because of this property, parallel activities by the parties are possible and emerge that must be coordinated with the activity of engaging in dialogue. However, it is not possible to engage in parallel activities and treat the other's contribution as a transient performance. Thus in the example both pairs of coparticipants are engaged in parallel activities but after K notices the construction of a possible contribution by E and R then, and only then, the visible performance becomes relevant as a contribution-in-progress to be addressed.

Example B3 in Appendix B shows the full complexity of the transcription notation adopted in the investigation. It illustrates the inaccessibility to the other party of the situated work of coparticipants. J and R are collaboratively engaged in resolving just what it is they are talking about that bears on the construction of a next contribution in the dialogue. However, this activity is unavailable to the other participants. Mutual access is restricted to the point where mutuality is problematic. Participants can observe some events, that are routinely available in copresence, and other events are completely unobservable. For example, the routine monitoring of contributions is possible but the continual, precise attention to each other's contributions in transient moments of production characteristic of copresence has been lost[2]. Speaking and listening are contingently achieved in talk, but in this virtual dialogue space they are practically unattainable given the skills and resources available[3].

The modality of the dialogue example in Grosz & Sidner (1986) was designed to be restrictive. The presumption is that the resulting dialogue will in essence be the same as everyday talk but without the interruptions and overlaps that litter that activity. Thus implicit in their methodology is the assumption that the circumstantial details of dialogue activity are not relevant to the work of intention recognition and coordination that they claim underlies and gives dialogue its coherence. This has been shown theoretically in Chapter 2 and empirically above not to be the case. Moreover, the intelligibility of the represented dialogue is taken for granted as is the reading of the representation. However, there is a mismatch. The intelligibility is to be understood in terms of a reading that relies on an inapplicable model,

2 An orientation to turns would be guaranteed by the use of an explicit set of signalling rules by the participants or that are built into the computer-mediation system itself. This is how legal courts operate to organise a routine order of examination, but it is not how conversation works. Also, it may be possible for competent touch typists to achieve a semblance of mutual engagement, lodging them in shared space of dialogue by embodied attention to a single locus of evolving interactive engagement.

3 They may be unattainable but then they are not a goal of the activity, nor does it result in a degradation of the quality of dialogue.

viz. the interchange model. The activity engaged in was no doubt intelligible for the participants, but the representational record must be justified as adequately rendering the details attended to by the participants themselves. The achievement of that intelligibility must itself be explicated and an initial step towards an adequate explication is to choose carefully the features that are important to transcribe. This section demonstrates what those features are in this case and what is misleading if they are not documented. Empirical methodologies for computer models can irrevocably miss the situated practices that give observable details their sense and rely on a biased reading of the reified record 'as if like conversation'.

It has been demonstrated in this section that careful attention must be paid to the practices of participants in constituting the sense of inscriptions in the VDS. Computer models of dialogue have not regarded the interactive circumstances as relevant to an underlying account of the organisation and procedures of dialogue. The interesting issue in this chapter is how the participants construct the dialogue with social-historical conventions, like conversational structures or practices, and in situ with emergent resources and circumstances, like adjacency in time and nearness in space. These are demonstrably endogenous to conduct and a constitutive domain of dialogue. They are unavoidable, useful, and essential.

5.3 Interactivity and Dialogue

Given the phenomena located in the last section that are unaccounted for, and as resources for situated action are replaced by an elaborate model, some concepts will be developed to form the basis for the analysis in the remainder of the chapter.

Ideas about interactivity are usually implicit within a theory or way of talking, because the meaning of the word is so obvious, yet diffuse and with many applications. It depends on the activity in question, both physical and social: reading a book, walking down a street, using a computer, or talking to someone. A person may inter-act within the physical world in terms of action-reaction events, or interact with another in terms of meaningful action by both parties. Typically, with a computer system this means allowing the user to participate in the computational process: with an interchange dialogue system instead of written text material a user can interrupt the unit of, say, 'explanation' at explicit points, ie. an explanation is not given as a monolithic whole, but in segments tailored to the user's behaviour and to dialogic interchanges in ways that are predesigned. However, and this is the crucial point brought out in this and the next chapter, the participatory dialogue itself is not interactive, ie. the organisation of participation is not interactively constructed by the participants themselves so that it may be impossible to collaboratively construct, interrupt or influence another's participation in the dialogue activity in situ.

Interaction can refer to the provision of an active role in the construction of intelligibility or action in an activity that is normally taken to be solitary and passive. For example, in fiction stories a reader interprets the story in a passive role in which it is not possible to change or influence the textual material on which the interpretation is founded. But in interactive fiction, a willing reader can actively influence the ordering and the substance of the material of the story itself that is normally considered to be static, eg. change the character or plot developments. It is acknowledged that computers may support a more active role for the user, thus participation of a person is sought for artistic, pragmatic, and efficacious reasons. However, in these cases the sense of interaction as mutual and reciprocal action by more than one active person is not to be found. The thesis is specifically interested in communicative interaction in which coordination and shared understanding is achieved by more than one participant, for example in dialogue.

A beginning can be made in the outline of an approach to this notion of interactivity by considering Goffman's system requirements of interaction, from Goffman (1981). These are theoretically descriptive and formulated in the manner of communication theory. For example, one requirement is that there exists a two-way capability for transceiving adequate and clear messages. Also, there must be backchannel capabilities for reception feedback, and contact, close-down, and turnover signals for turn-taking. But, this approach is much too close to communication theory. See Wilson et al (1984) who compare signal theory with the local opportunity theory in the study of conversation. Also, there is a similarity between the emphasis on coordination and recognition of markers and signals, and the coordination and recognition of communicative intent in computer models such as Grosz & Sidner (1986), eg. "we find that participants in explicit interchanges tend to alternate, or take turns, in engaging in explicit actions. This implies that a set of signals or markers will be needed by which turns are marked as beginning and ending, and by which the expectations of the participants as to who is to take up the next turn is indicated." (Kendon 1988, p. 33). We need to have a closer look at the local processes of interaction, and what better place to look than the detailed empirical work of CA on speech exchange systems.

5.3.1 Participation and Local Management

In the analysis of the achievement of dialogue it was recommended in Chapter 3 that dialogue be considered as a set of practices in which participants are mutually engaged in coordinating their communicative actions and achieving shared understanding. In order to explicate some of the details of the dialogue space, perspectives on interaction from the point of view of the coordination of actions will be outlined here because participants must coordinate their participation within the physical possibilities of reciprocal action. The question must be

asked: How does participation get organised in interaction within contexts of reduced copresence?

In the creation, negotiation and management of shared meaning, doings by actors are systematically relevant to each other. However, the coordination of actions and their boundaries may be determined by external forces or from within the interaction itself using the materials available at hand. An example of external force is the modality that makes it physically impossible for one actor to take a turn until the current turn holder has explicitly allocated the next 'speaker' by a conventional signal. Also, turn-taking may be institutionally provided for, external to the local circumstances of the interaction, eg. the 'over' convention for signalling the end of the turn in citizen-band radio operation. But, determination from within the interaction is operative in telephone conversations, for example, in which an 'open floor' of participation is possible so that actions and their boundaries must be locally achieved.

How is the possibility of simultaneous action significant for interaction and the participants? By organising actions contingently and locally, act descriptions and boundaries can be mutually constituted in situ under the pressures of interaction. The 'open' possibility of action means that participants acting at particular times have added significance in coordinating and achieving mutual intelligibility. For example, in interchange dialogue in which the current 'speaker' always allocates the next 'speaker', explicit linguistic work must be done in the explicitly allocated next turn to repair the sense of or display interpretation of the current contribution. The work of close mutual coordination of turn boundaries in conversation is a rich resource for constituting the intelligibility of talk. For instance, a turn transition at a possible completion point can display to the prior speaker that the recipient has understood something of the sense of the turn by responding at that particular point, or a head nod by a recipient during the construction of a turn can influence the trajectory of the turn. But in interchange dialogue this work is undone and the current 'speaker' must construct a complete turn without help, ie. completeness is not competed over nor mutually coordinated.

A fundamental order of participation is to be found in speech exchange systems. CA explicates the local management and organisation of the opportunity to speak so that 'one-floor' of participation is maintained with one speaker speaking at any one time. Here are the important concepts derived from CA studies described in Chapter 3. Similar ideas can be found in Laurel (1986) with respect to active-role technology. First, the participation of those taking part must be organised in some fashion. Mutual and simultaneous actions must be coordinated; they are not random events in a meaningless world. An **opportunity** for action is socially or physically organised and constrained, and in opportunities there are **possibilities** which can also be socially or physically constrained. For example, opportunities to participate are organised at transition relevance places in the construction

of a turn-at-talk while adjacency pairs sequentially implicate a space of possible responses. Interaction has a describable **grain size**, ie. the patterns of enabling opportunity for action. For example, the interchange model common in computer models of dialogue like Grosz & Sidner's (1986) has a large grain size of interaction because a participant can only act at explicit junctures granted by the current speaker. **Significance** is the potential effect on the whole interaction, ie. the dynamics of activity[4]. **Dynamics** are those orders and patterns of activity that emerge in interaction. They are not contained or represented within the individuals concerned but are properties which arise in and shape the subsequent course of the dialogue, either unwittingly or as accountable features of that dialogue. For example, a dynamic of turn-taking is that accountably one speaker speaks at a time. Also, an emergent dynamic of topical talk is that because of the linear stream of talk and the orientation to adjacency then topics can unwittingly get dropped, they pile up, and consequently, if it becomes relevant, they are harder to pick up again as they get further behind as one of many topics that could be reintroduced. What is important in contingent and practical interaction is that action must be organised locally and mutually within the emerging dynamics of the interaction. For example, interruption has to be achieved, coordinated, managed and 'pulled off' collaboratively. Even disagreement, nonsense, or chaos have to be coordinated and make sense.

5.3.2 The Interactional Context: Constraint and Resource

Focused human interaction is organised within contexts of copresence. **Copresence** is anchored in the perceptual and communicative modalities of the body. It is a continual embodied achievement of a shared physical and temporal space of mutual monitoring in which participants are "close enough to be perceived in whatever they are doing, including the experiencing of others, and close enough to be perceived in this sensing of being perceived." (Goffman 1963, p. 17). Thus, "actual conversation is always 'situated', always comes out of, and is part of, some real sets of circumstances of its participants" (Sacks et al, 1978, p. 10). **Context** is a cover term for the relevancies in interaction drawn upon to construct the interaction that

> *includes the physical environment of interaction but is not something merely 'in which' interaction occurs. Aspects of context, including the temporal order of*

4 Note that opportunity, possibility, and significance come to hand post hoc or in breakdown and rationalisation. They are analytic terms. Normally and routinely, a space of possibility and opportunity in practical action or 'going on' is not explicitly addressed in every case so that opportunities for action are encountered in the course of activity, not explored mentally or 'in the abstract'.

gestures and talk, are routinely drawn upon by actors in constituting communication. (Giddens 1984, p. 71)

The context of interaction is in some degree shaped and organised as an integral part of that interaction as a communicative encounter. The reflexive monitoring of conduct in interaction involves the routine drawing upon of physical, social and temporal context in the sustaining of accountability; but the drawing upon of context at the same time recreates these elements as contextual relevances. (Giddens 1979, p. 83-4)

The reproduction of interactional context forms physical and social constraints on action from the **circumstances** of time-space to normative sanctions. In addition, an **interactional event** is an event endogenous and emergent to the interaction that is used as a resource by the participants, eg. internal pace or rhythm constituted through the interaction. But, it is important to see **constraint** and **resource** as figure and ground, ie. they are different perspectives on the same phenomena. Understanding constraint and resource can help a lot in explaining certain dynamics, eg. what might be understood as a constraint of eye gaze monitoring of signing activity in the sign language conversation of the deaf is also exploited as a positive resource for managing turn allocation because signers can avoid mutual gaze so as to keep a turn, for example.

To illustrate the importance of constraint and resource, take the physical writing pen and paper versus the computer mouse and window technology common on personal computers. It is common for users, working in a multiple window workspace and having to coordinate the current visual workspace with the placement of current activity, to act in the wrong window, eg. a user typing a memo in a programming window. Why does mis-windowing occur? A simple explanation is that in writing on multiple sheets of paper, the actual physical writing process is very much a constraint on the activities of writing on a particular sheet, ie. the sheet and pen must be in contact through the bodily activity of writing. This is also a resource for the coordination of the writing activity with the next action's placement, so it is not easy to write on the wrong sheet given that the bodily actions of writing are so closely connected with the decision to write on a particular page. However, the removal of the bodily constraints, through using a keyboard that distances the writing process from the graphic medium, means that it becomes easy to misplace a set of keystrokes. The mouse now serves as the 'middle man' between the graphic medium and the activity so a user can easily forget where the typing will appear given the absence of a constraint tying 'inscribing' activity to graphic material. Reichman (1986) tries to explain the misplacement of actions by users in window systems through the metaphor of conversational embedding of context spaces in her model of dialogue. It is claimed that a user is engaged in dialogue with the interface and windows can be thought of as dialogue context spaces. Movements between them need to be explicitly signalled with mouse clicks which are comparable to markers like "anyway" in conversation.

Thus, a misplaced action is due to the user forgetting to change context spaces in the dialogue between the user and machine. Unfortunately, this explanation confuses conversational resources with action resources. Instead, the user is engaged in an activity in and through the interface. The interface provides resources for interleaving multiple activities, just as if at a workbench or desktop[5]. So, this example shows that what are constraints from one viewpoint are necessarily resources from another, and it is easy to misapply the 'dialogue partner' metaphor. To return to communicative activity, the focus on correctness has led to a similar one-sided view of certain phenomena and this is apparent in the view that self-repair in conversation, eg. hesitation, cutoffs, are signs in the speech stream of disfluency or a lack of competence or control. It is hard to sustain this view when the findings of conversation analysis are taken into account where it has been shown that these performance features of speech are positive resources for eliciting mutual gaze in face-to-face conversation (Goodwin 1981).

5.4 Analysis of the Virtual Dialogue Space

An interactional analysis of the achievement of the virtual dialogue space is undertaken next. The virtual dialogue space in the study is a computer-generated visual display space on which spatially distant parties can produce distinctive graphic objects as if both were copresent and typing on a physical sheet of paper in front of them. A participant can type graphic symbols within the constraints of the technology and conventions for spatial presentation that will appear simultaneously on both monitors in their separate locations. They can do this both at the same time in their respective windows or graphic spaces, ie. they cannot operate on the same visual space but on adjacent ones. The sense of this being a shared space for dialogue is a social achievement, ie. the relevance of visual elements to each other as contributions in a developing dialogue in the virtual space must be achieved continually by the participants.

Before presenting analyses of the VDS in this case it will be compared with some other similar systems that are available. The VDS is quite different from 'asynchronous' modalities, for example electronic mail. Variants of this popular modality are designed for temporally and spatially distanced correspondence. Participation is organised around the computer-based interchange of discrete textual documents over comparatively long periods among multiple participants. Because of the extremely restricted access to mutual circumstances, the

5 That circumstances are an essential resource for solitary activity using artifacts has recently been acknowledged by some psychologists and designers, eg. Norman(1988).

significance of time is ambiguous. COSMOS (1988) have argued that because of the asynchronies between times of creation, transmission, receipt, and response, a sender cannot find the sense of time passing without anything happening significant because it is unclear whether the recipient has received the message, received it but has forgotten to reply or is in the process of replying, received it and doesn't want to reply, etc.. These properties effect the communication that is possible and are fundamental to the nature of asynchronous computer-mediated communication.

Many 'synchronous' computer-mediated modalities are available similar to that in the study, but with slightly different characteristics and varying degrees of mutual access. For example, the STARLINK system allows that each user can construct a message in isolation. It is then explicitly transmitted to the other participant, whereupon it is received as a complete message. When it is sent, the message interrupts the other participant's message construction, and then allows that person to continue with the now interrupted message. So, there is an asynchrony between the creation and transmission of messages, but receipt is almost guaranteed because of the intrusive appearance for the other party. Participants can organise their actions as interchanges of turns - sequential message passing - or they can exploit the interruptive character of the modality. The latter case results in a slightly disturbed sequence structure and the possibility for a party to redesign the contribution on the basis of the interrupting contribution. The possibilities for the interrupted participant revolve around the constraints of the modality. The message can be completed as planned and transmitted; the message is abandoned and a new message constructed; the message is aligned to the accommodate the interruption, and sent; or the message is completed and new material, explicitly orienting to the interruption, is worked on, and both are sent. Because of the asynchrony between creation and transmission and the interruptive character of message transmission it is likely that time can be spent addressing more than one issue and interruptions lead to new topics being addressed. Thus, multiple topic threads are possible in this modality. This modality is like a form of electronic mail correspondence, but the process of construction is regularly interruptable by the transmission of messages by another party.

Another interesting modality in everyday use is the TDD telephone aid for the hearing impaired. It is similar to WRITE, which is a simple synchronous communication modality available on multiuser computer facilities. Both parties are spatially distanced, but they can use telephone network to communicate. One party dials the other and after an appropriate summons signal the other party responds. Both use a keyboard plus visual display interface to communicate by typed messages. The keyboard is a standard QWERTY, and the display is 40 characters long and scrolls from right to left - as you type, characters file along from

right to left and disappear off the display at the far left. In its most basic everyday use, participants alternate turns at typing a message, which appear on both displays simultaneously. There are facilities for recording a dialogue and replaying it, and for printing a hard copy. Also, a user can compose and edit a message prior to the actual call and send it at high speed during the call, thus reducing cost. A set of conventions is used for controlling turn taking as the display can be accessed by both participants at any time, ie. it is a simple shared visual space. Without conventions, garbled messages would be produced because the participants would inter-mix their individually coherent words made up of typed symbols resulting in a stream of unintelligible nonsense. The conventions and associations between particular symbol forms and possible interpretations are given below - they are thought of as strict rules to follow, as reported by a user.

GA - short for "go ahead".
Q-like shape - short for "question".
GA OR SK - short for "go ahead or stop keying".
SK - "stop keying".
SKSK - termination.

The conventionalisation of conversational resources is apparent in the provision of explicit means of managing the turn-taking. Turns are current 'speaker' controlled and are transferred at explicit points. Also, "GA OR SK" provides a pre-closing resource. An example of a sample dialogue from the manual is given below:

```
DAD: GOOD AFTERNOON  JOHN HERE GA
SUE: HI DAD  SUE HERE  PLS CAN I STAY AT THE
     LIBRARY TIL 5 Q GA
DAD: THATS FINE THX FOR LETTING ME KNOW SEE
     U THEN  BYE NOW  GA OR SK
SUE: BYE DAD  SKSK
```

The VDS in this study is interesting in that it allows for simultaneous construction of contributions to the dialogue and also monitoring of that construction that is unavailable in simpler modalities. Creation and transmission are not asynchronous, and receipt is not necessarily asynchronous. The monitoring is disembodied and only available through the visual space of the dialogue activity because participants are not copresent.

Before embarking on the analysis of the VDS, a word of warning given the brief descriptions above of the more restricted modalities. One cannot simply conflate the properties identified in studies of restricted settings of dialogue to get a full understanding of the holistic nature of copresent talk. Activity in copresence offers emergent resources for engaging in that activity

that are dropped out of accounts of the properties of restricted modalities. For example, the creation, transmission and reception of messages are not simply the components of copresent talk that are undertaken synchronously. The participants' mutually observable actions are not just constructions that happen to be created, transmitted and received at the same time. This way of thinking means that a mutual process is reduced to a manipulation of a product from speaker to hearer, but speaking and listening in copresent talk is a continual routine accomplishment within the circumstances of copresence. In restricted modalities of presence, features emerge as distinctive phases with identifiable social or physical constraints, eg. the asynchronies in electronic mail emerge as explicit non-features - an unachievable and unaddressable synchrony - that are fundamental to the nature of the modality. These features should not be reified as explicit features of copresent talk nor as the only components of that activity.

In contrast to the modalities described above, the VDS is an 'open-floor' of possible, synchronously monitorable action, ie. a participant has no physical restriction on opportunities for acting. But the relevance of actions as contributions in a developing focused interaction is a skilled achievement, so in this special modality the construction of intelligible contributions to the evolving interaction must be achieved socially in the local contexts and circumstances of their conduct. Because of the peculiar and particular properties of this modality, new organisations emerge through the participants' sense making practices partly borrowed from talk. The study will particularly focus on the following topics:

* The appearance of the virtual space to the participants and the possibilities and opportunities for action in that space given its circumstantial constraints. Because of the peculiarities of this space, features of text and speech are present and differentially orientable to.
* The visual and temporal monitoring of the construction of contributions.
* The weak mutual access that participants have to circumstances of situated action, and thus the emergence of parallel activities.
* The local management of participation in the virtual dialogue space with an 'open-floor' of opportunities. Complications and breakdown arise in the interaction as the participants manage their activities through the circumstantial resources. For example, the use of turn taking resources in managing participation is considerably weakened in this modality. Because of the physical properties of the graphic space, the coordination of communicative actions is much looser. For example, the 'single-floor' - or arena of communicative action - is not such a valuable commodity as in talk, because the graphic trace releases the participants from continuous mutual monitoring.

* The relevance of contributions to a developing dialogue. Adjacency and sequence are still valuable resources for constructing and interpreting contributions.
* Particular phenomena emerge in the activity. For example, the phenomenon of 'double dialogues' is found in fast dialogue, where the temporal singular stream of adjacent actions found in talk is not so routine. Some notion of adjacency is still applicable, but, because of the graphic trace of prior actions and the lack of mutual monitoring, two chains of dialogue can develop. Both are currently under construction by the participants, but at different times, and adjacency in each chain is still an important interpretative resource.

5.4.1 Constructing Action

This section will explicate how the participants worked in and through the technology for constructing contributions. The virtual dialogue space is a textual surface on which they can create and edit visual symbols within the space granted by the graphic display and keyboard input technology. The technology is designed with certain general goals and has specific behaviours, however, the user can discover, play, and abuse it in the course of their activities. Thus a technical description of performance characteristics does not give an understanding of how the technology came to be used in particular ways. The technology is easy to use as such but its routine use relies on certain skills and practical experience. We will find in this section that the technology becomes a constraint and a resource for constructing and interpreting contributions.

The VDS is designed to allow two spatially remote participants to communicate through a mutually accessible visual work screen on which typed characters are displayed. The communication link is engineered to be usable by people, and no instructions are given except a brief manual entry. The design is such that a naive computer user should be able to communicate using the link provided they can use a keyboard and observe the screen. Because of the possibility for simultaneous typing, participation must be socially managed. This is much more versatile than other interchange modalities.

The VDS (see Figure 5-1 in Section 5.2.1) is a visual space containing two separate graphic spaces which are 8 characters deep by 39 characters wide. Participants can press buttons on the keyboard and the appropriate character will be displayed, immediately upon key pressure, in the appropriate window on both monitors. The small screen width encourages short, quick responses. Activity in the window is managed through the keyboard and the observable opportunities provided by the display. Placement of contributions is restricted to the locus of cursor movement as participants type. Appropriate key presses will produce

visible characters in the current cursor position whereupon the cursor will move one place to the right. Only the RETURN key can influence this trajectory by moving the cursor down one line and to the beginning of that new line. Thus the contributions must be constructed as a linear string of characters stacked from left to right. Planning ahead may help but the activity is fundamentally prone to unforeseeable problems. This leads to recovery practices that attempt to salvage or repair constructions and interpretations which emerge in the course of constructing. The cursor and the lowest positioned line in the window are resources for determining the state of the dialogue activity, eg. what the last contribution was, where to type the next character, where an expected contribution by the other participant will appear.

The ownership of a window could be routinely discovered during the developing dialogue by inspecting the spaces for identifying textual items, or by typing and observing the location of the new characters. When a participant types the characters appear in their allocated window and the cursor remains in the next possible character slot in that same window. Only if the other participant starts typing do characters appear in the other window and the cursor remain in that window. However, for coparticipants H and L in Example 2 below, the relation between the current cursor position and the allocation of windows becomes problematic. They have just started and are unfamiliar with the VDS. The other participant has typed "hello" and both H and L recognise this greeting contribution and that a response is expected. But towards the end of the suggestion of a possible candidate for a response "type in hello \ba:ck perhaps", L notices that they are not in the right place for giving a response "i thought we were up there?". The expectation revealed by this fragment is that their window for responding is the upper one and that the cursor position indicates the next typing place. The cursor is in the lower window and thus they are not in the right place and access to what they see as their window is now a problem for them. This is reasonable as the normal operation of the cursor during typing is in the next positioned slot for a character on the right in that same window, ie. in this case, the other party's. Rather than discovering the automatic operation of the cursor, ie. it 'jumps' to the correct window upon typing, they attempt to resolve the 'problem' based on their first assumption. H suggests a way out of the dilemma. The cursor is wrongly positioned and they must explicitly move it themselves. "Use the cursor" focuses on the availability of the cursor to be manipulated as a spatial object in the graphic display. Normally, the allocation of windows and access to those windows is not a problem, though the significance of items in them is problematic.

A standard QWERTY keyboard is used. Characters appear at the current cursor position by pressing the appropriate key. Some key presses do not give a visual character - they are impotent for the practical purposes of the activity. The full range of alphabetic characters and punctuation symbols are available. A range of other characters are available but were

not used. In addition, the keyboard can perform many operations on the display and the computer system itself on which the VDS is running. The DELETE key moves the cursor back to the left one place and blanks out the character in that position. Characters on previous

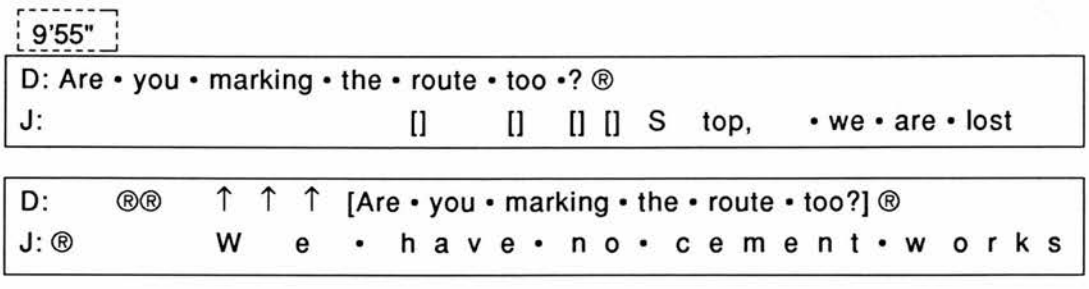
EXAMPLE 2 - [M9: 15"-20"]

```
      H:    well we could type in={hello}
      L:                              [{hello}
      H:    right- write- type in hello \ba:ck perhaps.
→     L:                                        [i thought we were up there?=
                                              (( ps                        ))
      H:    =well jus- use the cursor.
```

lines cannot be deleted. This is discovered by some participants. The RETURN key moves the cursor down one line and to the beginning of that new line on the left of the screen. One set of special keys to the right side of the keyboard are used by one participant. These are line editing keys that allow a user to modify an already constructed line. For example, one experienced user of the keyboard-monitor technology used it to redo a contribution as requiring a relevant response (see Example 3). A and D construct a question: "Are you marking the route too?", then a contribution by the other party is found to be immediately relevant to the cooperative activity of route construction. This contribution, "STOP, we are lost", is not a reply to the question and D displays the absence of a response using the technology available at hand. D moves the cursor to the line containing the question "Are you marking the route, too?". That line has now been entered in the record and thus is available for inspection, so D uses the arrow cursor manipulation keys to reconstruct the same line "Are you marking the route too?" at the expected place where a current contribution could be constructed, ie. it appears again on the bottom line of the window underneath the original. Thus D accomplishes a reformulation of the prior question "Are you marking the route too?" by repeating it, and thus displays it as a question still to be answered that was not addressed by the intervening contribution "STOP, we are lost" by the other party.

Some key presses can perform operations that are not explicitly visible but are revealed in the course of the activity. In Example 4, L appears to lock the keyboard while constructing

EXAMPLE 3 - [M5: 10'9" - 10'19"]



a contribution. There are also many keys or key combinations that can have a function but they have no visible effect in the VDS. In Example 5, E has just mistyped and disturbed the construction of a contribution. The coparticipant suggests that E just type the contribution again and E reinitiates keyboard activity. But after three key presses no visible effect is found. E exclaims "o- oh-" as the inadequacy of the action is revealed.

The characters typed by participants are semi-permanent as inscriptions on the display. A participant can press keys in order to construct a line full of characters, but before reaching the end of the screen display, the RETURN key must be pressed in order to start a new line. For each press of the RETURN key the lines above move up one, the one that is currently

EXAMPLE 4 - [M9: 48"-55"]

```
      H:    .h right.  ask where the start is.
              (pause)
      L:    ®
              (pause)
      L:    ((types))
            ((looks to the screen, but no characters))
→     H:    oh (.)
            ((looks to keyboard and back to screen))
            °can't type°
      L:    We need to...
```

top and the oldest is removed from the display, and the cursor appears at the beginning of a fresh line at the bottom of the window. This is called scrolling. Thus a string of characters is entered on the display that becomes unchangeable, but still visible, when the RETURN key is pressed, so the participants have some visible record or trace of what happened in the prior dialogue. The extent of this trace is not defined temporally as having a fixed time limit before erasure of an item, though it does have a bounded history, but an emergent dynamic of the situated use of the window mechanism.

EXAMPLE 5 - [M8: 2'10"-2'20"]

```
      E:    oh
      R:     [(<don't worry)
      E:           [bugger it
      R:    JUs- (.) <jus' do it again. (.)
            type cavendish again.
      E:    ((types but nothing appears on screen))
→           ((finds a problem with keyboard activity)) o- oh ((tries again))
      E:    Cavendish ...
```

In Example 6, participants use the display as a resource for remembering the name of a street. D and M are engaged in locating a major road, called Hills road, that will help them locate the starting point on the street named Blinco Grove. They find the major road and D

reorientates them to the main task by saying "so we're looking for \blinco street." Just after the production of the main noun D shifts gaze to the screen followed by M. Just a few contributions ago the street name was stated and is thus visible in the record of inscriptions on the display. That adequate confirmation is found on the screen for D's remembering of the name is displayed in D's repeat of the name and by coparticipant M's "ok" as they both reorient to a search on the map surface. However, though agreement is temporarily reached over the street name it is in fact incorrect. The confusion is discussed a little later.

Through the process of communication, using the resources and constraints of this modality to constitute the significance of communicative action, a window on prior dialogue arises. The display contains particular sedimented records of what went on, but many significant features of the interaction are unrecoverable simply by inspection. Also, many features are only recoverable by the participants if they noted and can remember enough to reconstruct

EXAMPLE 6 - [M5: 4'10"-4'20"]

D&M are both looking on the map surface for a road.
D: HILLS ROAD.
M: where,
(pause)
D: hills road there.
M: [oh great.
M: right.
(longerpause)
→ D: so we're looking for \blinco// street.
→ ((looks to screen))
M: ((looks to screen))
D: \blinco //street.
((looks to map))
D: ((D looks to screen, then both to the map))
→ M: °ok.°

what happened. The characteristics of the keyboard and display thus constrain the construction of communicative actions. Also, the technology provides resources for developing or repairing the dialogue in situated ways. For example, the RETURN key commits a participant to an entry on the window record because it cannot be edited. However, the key itself also provides the means to remove the entry from the record. In the extract in Example 7, D constructs a contribution, "We have to guide" which is disattended while they attend to the other party's construction. It is then restarted by D and entered on the record by the return key but it is still incomplete.

The in-progress contribution has now been entered on the record as two lines making "We have to guide you round certain". Up until this point M has been orienting to, eg. out-louding,

the simultaneous ongoing construction of the other party. D begins to construct again on the current line which is immediately below that of the unfinished contribution. This could be a completer, eg. "places", or a response to the other party's contribution made relevant by M.

EXAMPLE 7 - [M5: 23"-58"]

```
    D:    We have to guide
            ((D and M attend to other party's construction))
    D:    you round certain ®
          p
    D:    ok
    D:    ((deletes the 'p' and tries to delete other lines above))
→         ® ® ® ® ® ® ® ® ® ® ® ®
→   D:          [i'll have to get rid of that
    D:    We have the...
```

After one character, "p", D looks to the screen and says "ok". This and his next actions acknowledge the immediate relevance of responding to the other party's construction. He deletes the letter and attempts to delete further characters. But the next possible character to be deleted is on the previous line if the trajectory of the cursor is retraced. That D wanted to delete the troublesome unfinished contribution is made clear by the use of several presses of the RETURN key itself, that gave rise to the problem, to remove the offending line and all other lines from the record. D comments, "i'll have to get rid of that", during this activity, and then proceeds to construct the appropriate response on the fresh display. Thus D has managed to skilfully extricate them from responsibility for a possibly ambiguous construction that could be taken as a completer of the half-finished construction or a valid response that leaves a half-constructed contribution expectably requiring completion later. This analysis illustrates that the RETURN key is not just for entering significant and intelligible items onto the record as a by product of starting on the next activity, but can also delete entries from the record that may be troublesome, misleading or superfluous.

Thus the operations provided by the interface are both constraints and resources for constructing dialogue. Functionality can be designed but also it can emerge in the situated practices of those trying to construct intelligible actions from the materials at hand, and the significance of an operation is only to be found in the situated circumstances of its use. The technology provides a means of manipulating the space that both creates possibilities and opportunities for action in the activity and its use is shaped by the demands of that evolving activity. Another situated use of the operations provided by the interface is as follows in Example 8. The RETURN key is used to differentiate the types of contributions. Also, capital letters are used to highlight words and phrases. Single lines are placed between ongoing constructions that must be split over more than one line. But when a new contribution is

started two lines are placed between the prior and current, eg. when D responds to the

EXAMPLE 8 - [M5: 20'04"-20'22"]

```
    D:    continue along the footpath along to ®
→         CHERRY HINTON ROAD
    D&M:  ((Both look to screen))
          ((M reads a new contribution by other party, "STOP !!!!!!!!!!"))
    M:    {stop} (.) ur::.
           (pause)
→   D:    ®®OK
```

directive "STOP !!!!!!!!" with "OK". The screen display extract in Figure 5-1 in Section 5.2.1 also illustrates this usage. Note that D constructs contributions so that capitals differentiate the location noun phrases from the rest of the syntactic construction.

The coparticipants must construct tangible contributions from their situated work. A communicative response or act must be constructed by typing characters, and that requires producing an intelligible verbal string. Most constructions are mutually worked and agreed

EXAMPLE 9 - [M8: 3'02"-3'23"]

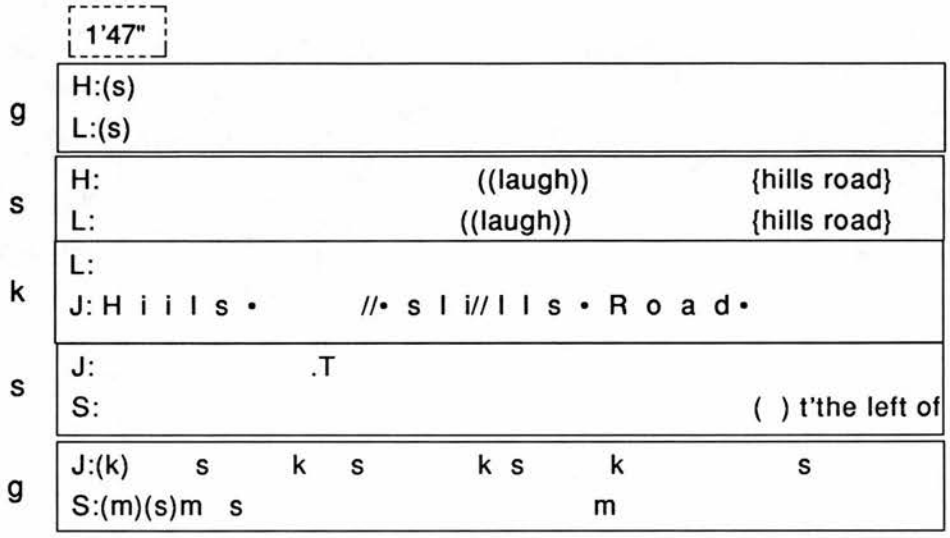
```
        K and G think they have found the starting point.
        G:    ok i'll put um, ((looks to screen))
              south  ((points to screen and K looks)) south west of brookfields district.
              ((both look to map))
              ((K points loudly))
        K:    yeah
        G:    yes ((looks to screen)) right ((looks to keyboard))
→             located i'll put  right? ((looks to screen))
        K:    ok.=
→   G:       =®school located®
```

upon. However, the resulting constructions are not simple reductions of speech. The coparticipants' talk about constructions does not specify their exact typed form but requires work to be made relevant in the immediate dialogue circumstances. Agreement can be reached over a possible current contribution, yet the final form is different (besides the contingencies of the other party's actions). The next fragment in Example 9 illustrates this. The typist G orients to the construction of a contribution given that they appear to have found the starting point. After a brief digression, that confirms the location, G suggests a possible informing contribution "located" for discussion and agreement. K agrees and G immediately starts the construction. However, G types "school located".

In the chronic process of production of action, simple errors are made or new problems arise that require a change in the current string. For instance, there is an orientation to correct spelling, for the sake of correctness or in case of ambiguity where correctness is necessary.

The DELETE key can remove from the screen items in the current line, but cannot remove items that have already been entered and are visible on the semi-permanent record, ie. in a fairly recent period of time. The use of DELETE key to remove items is illustrated in Example

EXAMPLE 10 - [M9: 1'40"-1'55"]

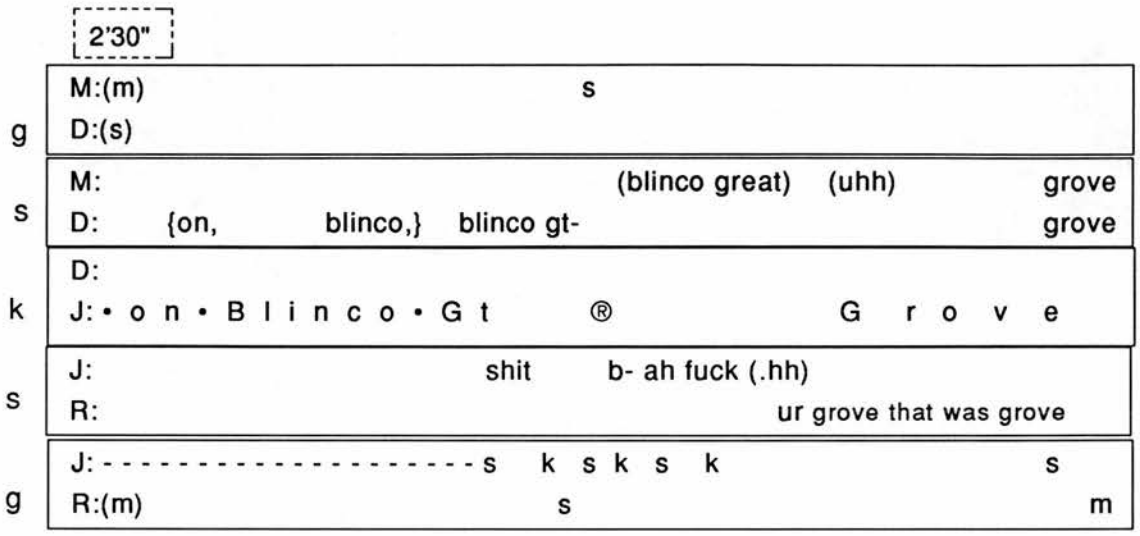


10. J types "Hiils ", notices the spelling mistake, tuts, and repairs by deleting back to the first "i" and typing "lls". The other party recognise this activity as repair initiation and thus now see the original form as a mistake and laugh. The correct form is recognised immediately after the retyping as "Hills road". Note that the repair activity is monitored by J's coparticipant. S looks to the screen just before the typo-error and looks away only after the correct form has been typed.

A different example of situated self-editing is to be found in the following extract in Example 11. J is just finishing a contribution when he makes a mistake and exclaims "shit". The location should have been "Blinco Grove" but J types a letter 't' instead of an 'r'. D attempts to delete the typo-error but mistypes a RETURN, exclaiming "ah fuck". Now the error is undeletable using the DELETE key. However, J types the street-type again and this is oriented to much like a restart that self-repairs in conversation. The problem with the typo-error is also demonstrated by the recipient party. They try to construct an interpretation of "Gt", eg. "gt-" or "great". The contingent construction of a repair item on the next line by J is successfully found to be a replacement of the item "Gt", and not the beginning of a new contribution. It has already been shown that reading the sedimented record is problematic for an analyst. It must also be so that it is not clear from the record what happened for the participants themselves at a later time. In an earlier example (Example 6) the display was shown to be a resource for remembering past dialogue contributions. However, in that

example the remembered noun phrase was misread from the display as "Blinco St". This is understandable because "Gt" does look a little like "St" at a glance. Thus the recovery of old and apparently already known information by inspecting records at the present time is a

EXAMPLE 11 - [M5: 2'30"-2'48"]



contingent practice.

From the physical and temporal engagement in constructing intelligible contributions as actions in a dialogue, and interaction between the parties, arises the particular characteristics of this modality. The structure of the communicative encounter emerges by the local work of both parties in situ. For the participants windows are continually available for constructing contributions-in-progress, and for editing and deleting them. Contributions are constructed and found within the achieved order of the visual space. Intelligible dialogue must be constructed within the evolving range of opportunities and possibilities provided for by the VDS in the moment of construction. But the mutual achievement of the intelligibility of the inscriptions as dialogue contributions requires explanation.

5.4.2 Visual and Temporal Monitoring

This section will look at the perceptual modalities which participants can draw upon in constructing their communicative actions or monitoring those of the other party in the dialogue. Reflexive monitoring has already been mentioned to some extent in the last section, eg. keyboard typing and reflexive monitoring of the graphic window. Monitoring of each other's communicative actions is the main topic of this section.

Participants have very limited access to the activities of the other party. There is no bodily or spatial copresence. The participants only have mutual access to the graphic space on the

monitor, where communicative action takes place. Because contributions are constructed visibly on both monitors, some elements of the chronic process of production of communicative action is available for monitoring, eg. a recipient can monitor the trace of the typing process of the producer. The only resources for monitoring the activities of the other party are to be found in activity because there is no background of observable inactivity, eg. a landscape in which actions are produced. Participants can be engaged in communicative activity when typing, but periods of no-typing leave the other party with no perceptual evidence of what else the other party could be engaged in.

Example 12 illustrates a weakness of other-monitoring. During the simultaneous contribution

EXAMPLE 12 - [M5: 1'20"-1'52"]

```
      D:    Is the A1303 in the top left hand corner
→     J:          [O u r  s t a r t i n g  point is
      ((D and M notice the uncompleted construction by J.))
      M:    {our starting point is:}
              (longpause)
      M:    what?
              (longpause)
→     D:    go on !
              (pause)
      J:    yes
      D:    yes  good that means
```

constructions, D addresses the construction by J, and thus moves to a more mutual focus that makes the uncompleted construction relevant. M reads the construction into the talk demonstrating its incompletion with an extension of the terminal syllable, and after a long pause, a quizzical "what?". After another long pause, D orients to the noticeably absent completion by the other party. There is no possible means of monitoring the reasons for the inactivity, so D initiates repair with "go on !" It is not possible for participants in the VDS to address the activities engaged in outside of the VDS by the other party; they only see the absence of the expected VDS, or else read of an activity reported later in the dialogue. There is also a similar problem on the telephone. A participant can notice a period of absence of speech activity when expected, but it is sometimes difficult to ascribe just what is the reason or cause of the inattention. The soundscape then becomes important not as a place to speak but as a 'window' on a possible world of sounds explaining what is happening at the other end, eg. difficult breathing, other voices, etc..

The physical characteristics of transience and permanence of a medium are important in this case. Speech and the signs in sign language conversation are transient, though they can be recorded and reviewed, whereas texts are generally permanent and imply a distance

between producer and recipient. Because of this property, the systematic synchronous coordination of production and recipiency is not necessary. If the graphic display was transient, as on a graphics-decay terminal, or the production was transient, as in rapidly decaying free hand writing on a graphics-decay terminal, then synchronicity would be essential.

Let us consider for a moment the differences between talking and writing. Interactional events emerge in transient dialogue and other monitoring leads to local construction of transient conduct, where the reflexive production of action has transient characteristics that are significant. This is not the case in writing texts because the editing, physical writing process, etc. are part of the chronic process of production of that text, but are not available to the reader except of traces in the sedimented product or cultural object. But, in communicative interaction, the processes of production are integral to the action, eg. the temporal dimension in the flow of speech conduct, and also to the coordination of actions in interaction, eg. pausing has differential significance for turn-taking in conversation. Equally, the process of production can be 'interrupted' which is unlikely or impossible in writing. Therefore, the process is contingent and subject to occasional recipient design. Monitoring of the other participant's actions may lead to editing, or drastic abandoning, of the current string because of problems arising through re-interpretation or other-initiation of repair.

As a small digression, Example 13 is an interesting illustration of recipient activities in which a dialogue contribution and the coparticipant's route construction do merge successfully so that the two activities of constructing contributions and engaging in mutual dialogue are not

EXAMPLE 13 - [M5: 6'50"-7']

J has is just finishing typing the contribution "what is the first destination?"
and M and D are working on the map surface.

M: why? why not go round like this? ((gesturing on map))
D: good move.
(.)
D: brilliant.
M: [right,]
(.)
M: an'then back.
D: [br i l l iant.]
M: <So we want to head ((D looks to screen)) up to here. ((M points on map))
D: [ok,
→ D: what is the first
((D looks back to map and M moves closer to map surface, repointing))
destination.
→ M: [the first one seems
to be this hospital, but without the name.
D: [ok [.h [okay
((D looks to keyboard and starts typing "go north"))

at odds.. Initially D and M are both working together on the map surface in order to construct an appropriate overall route. M suggests the first place to head to given their global plan, "So we want to head up to here", but D looks towards the screen whereupon he finds and reads a new contribution by the other party. However, M is still working on the map planning a route to the first destination. The out-louding by D is interpreted by M as a relevant question by D in their mutual activity. Thus the other party's question smoothly finds a relevant place in their parallel activities. Rather than coparticipants finding the other party's contributions as disruptions or digressions to their current activity, in this example the contribution is read into their activity at an exactly appropriate point.

Because of the permanence of the medium and the possibility of the other party monitoring the process of production, there is a tension between the talk-like process management of the interaction and the writing-like chronic process of monitored production without reception monitoring. That is, contribution construction is a process subject to the pressures of interaction and its construction as a communicative product. There is a tension for the 'producer' and 'recipient' between the differential pressures of monitored and unmonitored production, eg. between the close monitoring of transient production and the distanced 'when I am ready' interpretation of already produced communication, where its significance as a prior is reconstructed from the display. This means that the participant's conduct may best be described as forming a continuum between orienting towards producing a stream of communicative action to be monitored and a string of characters to be read at some near future point.

The VDS can lie awkwardly at points between these two because participants can address contributions as 'nows' or return to them or find them later as 'priors'. That is, a contribution can be produced as a complete unit to be received in a future time, or it can be constructed in monitored presence as one in progress. It serves as both constraint and resource. This is important for interactive artifacts where the physical characteristics can be creatively designed and may lead to new or troublesome dynamics.

5.4.3 Weak Mutual Monitoring

There is no systematic means of achieving mutual monitoring in the VDS. A participant cannot sense or monitor the monitoring by another participant. Performance of a construction can be oriented to by a recipient as shown in Examples 10, 11 and 12, but construction cannot systematically display orientation to that monitoring - thus copresence is difficult to achieve. For example, a producer has very few resources for monitoring recipiency during production, except through the assumption that no typing activity by the other party implies recipiency. If a recipient is typing then this indicates that the other party is probably not monitoring

concurrently. There is thus an asymmetry in producer and recipient monitoring, and so 'speaker' and 'listener' synchronicity is difficult. This means that reception can only be ratified in the next 'response', rather than routinely achieved through the coordination of synchronicity essential with a transient medium such as speech. In telephone conversation there are no resources external to talk itself. mutual monitoring is a routine achievement in which recipient inattention becomes explicitly evident because of the transient speech and fine coordination of speaking and listening, eg. monitorable with backchannels and the rhythm and pace of turn-taking.

Production and reception can be distanced because of the permanence of the graphic medium. This is not possible in everyday talk, where speaking and hearing must be coordinated synchronously. Other monitoring of current production of actions is possible, but in contrast to conversation, mutual monitoring is not found in the data. But if expert users of the keyboard - touch typists - could monitor two visual spaces of production, then there could be a systematic orientation to a 'floor' of action involving speakerhood and some mutual monitoring. Maybe this could also be achieved through the provision of background monitoring, that would allow speaker monitoring of recipiency. This is the only thing that stops the speaker's linear production of a response moving towards a serial stream of speaker/listener conduct, where the ongoing process of conduct is a resource for the parties. Will this lead to a different organisation of interaction with expert use? That is, if the participants can monitor each other's typing and type at the same time, then the concept of an economy becomes important. So, the parties could organise their participation in a similar manner to transient conversation, based around 'one speaker/no gap', or based on optimising production time but limiting local sensitivity. This hypothesis could not be tested further using the corpus here.

There are many parallel or simultaneous activities during the course of the dialogue because of the characteristics of the medium - permanence - and temporal duration of construction, and the further lack of mutual monitoring. Parties can address an activity relevant to the dialogue or task while dialogue contributions in view, or other activities out of view, are being constructed. These parallel activities are normally consecutively ordered activities, by necessity in copresent talk, but are now carried out simultaneously. Simultaneous activities usually need to be coordinated in copresence as not having a claim on all parties' focused attention and need continual routine monitoring, but in the VDS parties can engage in activities that are unobservable to the other party.

5.4.4 Local Management of Participation

The previous section explicated how participants construct their actions and can monitor their own and others constructions in the VDS. This section will address the management of participation in the dialogue which forms the basis for the achievement of the relevance of those constructions as contributions to a developing dialogue.

It has been shown that participants are engaged in constructing symbol strings in and through the technology available. By monitoring the construction of these strings participants are engaged in achieving a virtual dialogue space in which constructions are built and interpreted as communicative contributions. Symbol strings in the space of the shared screen are systematically relevant as contributions by each party. But, because of the permanence characteristic of the medium and the lack of mutual monitoring, participation in the dialogue is not organised like normal everyday speech exchange. Normal focused talk in copresence or on the telephone is demonstrably organised on the basis of the turn-taking systematics discussed in Chapter 3. The local management of the turn-taking is such as to orientate to there being one speaker, no gap, and one floor of speaking[6]. However, in the VDS this is demonstrably not the case because there is a much weaker orientation without a focused 'floor' of opportunity in which participants constitute themselves as speakers who allocate and construct 'turns', and as listeners.

Participants must find sense in each other's constructions as addressing aspects of prior or current constructions and this requires they organise their participation in some way. It is physically possible for them to type at any and the same time, so there is no guaranteed physical or institutional constraint that supplies a strict external ordering on the constructions, eg. strict alternation of turns at typing. Instead, they must locally construct their participation such that the construction of strings of characters are significant as intelligible contributions in a dialogue. Participants do orientate occasionally to the allocation of 'turns' but the nature of the VDS radically transforms the organisation of participation. Rather, participants are orienting to the construction and interpretation of 'contributions' to the dialogue. The carefully coordinated activities of 'speaking' and 'listening' are not necessarily primary activities in the VDS because there is weak mutual monitoring, and thus, mutual ratification of speaking and listening is problematic. Consequently, participants display orientation to the 'producer' and

6 Levinson (1983, p. 301) mentions evidence for the physiological and auditory possibility for simultaneous speech in Miller (1963). There does not seem to be any compelling constraint such that spoken conversation could not be organised otherwise, eg. involving simultaneous speech.

EXAMPLE 14 - [M6: 17'08"-22'20"]

```
     A:   OK?
     J:   yes
             (pause)
     A:   continue round to the A1134,
          or Loog Rd., and turn left.
          OK ?
     J:   do you mean alnog the footpath
             (longpause)
1    A:   yes, (pause) and turning into (pause)
1    J:                      [whe     r]
     A:   robinson way OK?
             (pause)
2    J:   i have no robinson way on my map
2    A:                             [we want to get to the 6th form
          college (pause) on LONG RD
             (pause)
     J:   is that Perse School for boys?
3         to get there I have to cut across (pause)
3    A:   its opposite
3    J:             [the park or es
3    A:                        [sedler taylor rd
3    J:                                      [else go a lnog way
          round
             (pause)
          ok I see the college now, do I
          cut across the park
             (pause)
          ?
             (pause)
     A:   there is a Rd marked on my map
             (pause)
          so that you will not actually
4         need to cut acroos
4    J:   [is the road west] of Addenbrokes
     A:   across.
             (pause)
     J:   ?
             (pause)
     A:   yes
     J:   does it go past...
```

'recipient' of contributions. Note, it is not that participants forget about turn-taking and disintegrate into random participation without order. It must be remembered that participation is a continual, routine achievement constituted out of the materials at hand. So, if new circumstances and contexts arise then participation will no doubt change accordingly, and in this case it does. A comparison between conversation and the VDS will be given next. It is based loosely around the observable phenomena of turn-taking in conversation described in

Sacks et al (1978) and using analyses of the VDS from the corpus. It shows that many phenomena to be observed in conversation can often occur in the VDS, eg. that speaker change recurs is the same, but also that they can be markedly different from each other, eg. that there is overwhemingly one speaker at a time in conversation is more relaxed in the VDS, and the no-gap transitions between speakers are not common in the VDS unlike everyday conversation.

1. Speaker change recurs, or, at least occurs (in conversation).

This is the same for the VDS. Example 14 shows an extended simplified example of a VDS dialogue. It is clear that participants A and J do recurrently contribute to the evolving dialogue and at first glance it seems quite like turn-based speech dialogue. It was chosen because it is quite similar though we will shall see that this is not the case in general. Examples of the beginnings of the disintegration of turn-taking can be found in this example. Ambiguous completion is shown at point 1 in which A answers "yes" to a prior by J and there is a pause. A continues at approximately the same time as J starts a turn and two competing analyses of that pause are displayed, eg. A orients to the pause as a gap occurring in the construction of an as yet uncompleted construction whereas J orients to the pause as a transition relevance point in the dialogue at which J can self-select. Thus the completion of that contribution by A adding to "yes" is empirically ambiguous. Also, early self-selects are to be found at points 2, 3, and 4. For example, at point 4, J 'interrupts' A's construction "so that you will not actually". J has apparently self-selected before an adequate transition point. The overlap that consequently ensues is resolved after four words by A dropping out. The extended overlap that occurs at point 3 is unusual in its pattern and that it is not attended to by either party dropping out. We shall see later that these two phenomena are mainly responsible for the emergence of quite different organisations of dialogue such as 'double dialogues'.

Two other examples of talk-based dialogue will be given in order to compare with the VDS. The first, Example 15, is an extract from the task conducted in the telephone modality and the second, Example 16, is an extract from the task conducted while participants were copresent or face-to-face. In the telephone modality participation was organised much like conversation. For example, long pauses are rare, overlap is brief, turn order is not fixed, and current turns address just prior turns in sequence. Also, the copresent modality is similar though misleadingly so because the participants were mutually engaged in coordinating non-verbal as well as verbal activities. Thus, the pauses are not signs of inactivity, just absences of talk activity. This is shown in the second extract, Example 16, where in the second pause J looks to W's map and W looks to J's map. The construction of the route is predicated on situated agreement and determination of an adequate location of referring

EXAMPLE 15 - [M2: 20-21]

R: and take the se:cond, (.) left, (.) ( ) shaftesbury road,
C: [right. yeah?
(1.0)
R: and go straight down there,
C: yeah?
R: and at the end of there (.) it goes straight on with, (.) what i assume is a foo:tpath or something,
C: oh i haven't got that marked (at all.)
R: [er::::]
(1.5)
R: it's a- (.) well no it must be a private road inside, (1.0) thee grounds.
C: Oh. (.) the GOVernment offices.
R: [ah : : : :]
(1.0)
R: yes. (.) you just go (.) straight on as if you, as if shaftesbury road was extended,
C: right. so, going back towards the railway track.
R: that's right.
C: yeah.
R: an:d'um, (1.0) keep going straight on=the road curves round,

items in their talk. A problematic referent, eg. they do not have the same one or it does not exist for one of them, is thrown up by their competing noun phrases "church lane" and "church end". J glosses the name by adding "or something" and pursues agreement with "yeah". But, both recognise a problem that requires them to check each other's map and this is what happens in the pause marked on the transcript record. Agreement with evidence is displayed in the next turn when J confirms an adequate understanding with "ok. right." The copresent example also leads into the complexity of talk and body coordination and relevance to the practical activities. This will be raised again in Chapter 6.

2. Overwhelmingly, one party talks at a time (in conversation).

Participants do sometimes orient to a 'floor' of participation in the VDS, but this is not highly valued. There are examples of explicit orientation in the coparticipants' talk and in the VDS itself to one party at a time. Some will be given below to show that the phenomena does exist to some extent. However, as we shall see later this 'rule' is continually undermined and simultaneous party talk is common.

In Example 17, M displays to her coparticipant that the other party has a right to a 'turn' which must be respected. They are waiting for a reply from the other party when D starts a new construction. This almost immediately overlaps with the onset of the other party's construction. M notices this and tells coparticipant D to "HANG ON hang on they're replying". M thus demonstrates in the talk an orientation to a 'correct' ordering of participation which D

EXAMPLE 16 - [M1: 4'25"-5'00"]

W: yeah there's a road there, an' you get onto rosemary, (pause)
rosemary lane?
J: [(alright)]
J: °nhn° =
W: an' onto church lane?
J: church end, or something=yeah.
→ (pause)
ok. right. ( )
W: terversham drift,
(pause)
J: ah: hh
(pause)
yea:h, that's=what is it ( ) that's, (.) that, (.)
cutting, (.) round there. (.) right.
(pause)
W: ok. yeah. Back, Back south again.
(pause)
an'then cuttin' back up north, ta church end t'the roundabout.
J: right. yeah.
W: ok.
J: a:nd then (swing)
W: [an'then a big loop, (pause) round t'the south.
(pause)
J: (n'thats) is that coming onto:,
W: (think) it's a roundabout ( )

has flouted. But D answers by claiming that contributions to the dialogue can be simultaneously constructed. M responds with "oh" indicating a change of state (Heritage 1984b). However, M is not convinced because a little later in the following dialogue in

EXAMPLE 17 - [M5: 22"-30"]

D: We have to
Others: [W e

→ M: HANG ON hang on they're replying.
→ D: [oh it's alright th' they can go together.
M: °oh°

Example 18 she sanctions M again. They both are monitoring the construction of a contribution by the other party when D displays a shift towards also constructing. M claims that this is inappropriate, "wait wait (pause) let them", as they should be allowed to finish the

EXAMPLE 18 - [M5: 31"-40"]

The other party is typing a construction. D and M are both monitoring.

D: yeah ((moves towards keyboard and gaze to keyboard))
→ M: wait wait
(pause)
let them
(pause)
→ D: no 'cause ((points at screen)) i type ( )
M: [oh right ok

construction uninterrupted[7]. D challenges this claim by arguing and providing evidence through talk and gesture that indicates the separation of the dialogue space into allocated windows that simultaneous typing is possible. M responds to this much more affirmatively and from then on the orientation to 'one party at a time' is undermined. In the first example, D does not continue the overlapping construction, but in the second he does which does not lead to any problems later except that J must accommodate the construction to the sense of the other party's simultaneous developing construction.

Another illustration of sanctioning, first off stage, ie. not in the VDS, and then explicitly in the VDS, is given in Example 19. The second explicit sanction leads temporarily to a return to 'one party, one floor'. These demonstrate that the notion of 'interruption' of a 'turn-at-typing' exists for some of the participants. J has just finished a projectably complete contribution, ending with "Cherry Hinton Hall" and continues with "Ang". But, D responds to the prior

EXAMPLE 19 - [M5: 19'16"-19'36"]

J: ...Cherry Hinton Hall
Ang
D: [The nusery is to the east of cherry hinton hall.
→ J: i wish this guy wouldn't interrupt when i'm typing (finished) ((laugh))

contribution thus unintentionally overlapping with J's 'continuing' construction, ie. in effect, they have both self-allocated and thus for J, "Ang" is a continuation, but for D the item does not exist yet. J makes a mistake and so looks up to the screen, notices the overlap and halts. At the completion of D's competing construction J mentions the incident to his coparticipant thus revealing his interpretation that the other party interrupted him.

7 Note that M's sanctioning of D's moves towards simultaneous construction are not claims that it is a physical impossibility, ie. that one at a time turn-taking is a property of the system.

A little later, J explicitly sanctions the other party with the contribution, "STOP INTERRUPTING" (see Example 20). J has apparently finished a projectably complete contribution ending "on our map". He pauses and then uses a '-' to demonstrate continuation. However, that pause is enough for D to self-select and thus competitive onset arises in which D legitimately responds but J can only find it as an interruption. The importance of a single

EXAMPLE 20 - [M5: 20'54"-21'05"]

```
     J:    The footpath is not marked on our
           map
             (pause)
     J:    -
     D:    [OK, r e t u r n  t o  C H E R R Y]
→    J:          [right come on .THH ( me) STOP interrupting (me )
     D:    HINTON ROAS
→    J:    [STOP INTERRUPTING]
     J:    Do we turn off...
```

focused 'floor' for J is demonstrated in the dramatic response that J gives in order to wrest the 'turn' back from the other party.

In Example 21, which illustrates orientation to some sort of 'floor' of participation, one party is interpreting the absence of a completion of a construction, while the other party find a completer based on the material already in play on the screen. J reads the construction as incomplete, ie. with an extended terminal syllable, and after a long pause suggests an ending. But, M is concerned about a problematic item mentioned in the prior dialogue. J does not respond and M poses a possible contribution dealing with that problematic item. J now claims that such a contribution is not immediately relevant at this point and would "put them off". J displays that it is the other party's 'turn' and they are in the middle of constructing a contribution. M doesn't continue with the suggestion, and they play 'wait and see'[8]. J constructs some other possible interpretations of the reasons for the absence of the other party's dialogue activity. Finally, the other party respond with a completer. This example is also interesting from the point of view of the other party as they attempt to find an ending for their construction. Rather than abandon it they work on finishing from the materials already in play. The incomplete construction has been entered on the VDS record and now an appropriate noun phrase must be found. S and R continually invoke the incomplete material as requiring an ending by typically repeating fragments of it in a quotation voice with terminal syllable extension. The possibility of repairing the material already in play is not entertained

8 Later, this query arises again and this time is instigated in the VDS.

EXAMPLE 21 - [M10: 11'48"-12'10"]

Others: ...which goes past the®

J and M try to interpret and predict the completion of this construction.

(longpause)
J: {which goes past thee:}
(longpause)
J: °railway line°?
M: [shall we query this jetty thing.
(° can't see that jetty°)
(pause)
(we just say we can't see a jetty on our map.)
→ J: um: (.) i think that will just- (.) put them off.
(longpause)
J: (i wonder what they're trying to do)
(longpause)
J: trying to decide what the footpath goes past.
(longpause)
Others: ((finally complete the construction with "cement works"))

The others, S and R, are engaged in finding a completer for the initial construction above.

→ S: {which goes past the:} (.) water (.) is it=
R: =<cement works.
S: cement works.
→ R: {past thee-}
(pause)
R: what's this stuff here.
...
→ S: follow the footpath ( ) {which goes PAST: THEE:}
R: past the jetty.
(pause)
→ S: {past,} (pause) hang on that might take them that way
y'see towards where it says jetty,
R: [oh yes (that's right.)
→ S: {which goes past the:}
R: that goes straight on. (.) i s'pose.
S: {that goes,}
→ ((S tries to remove the half completed construction, but it cannot be undone.))
R: no.
S: o'i can't bahck uph ((laugh)) .hh
→ R: {past the cement works} if you see that then ( )=
S: ={past the} cement works.
((S types "cement works"))

at this point. Unfortunately, an apposite ending is difficult to find and R suggests an alternative requiring the editing of some of the original material, eg. "that goes straight on" instead of "which goes past the". S repeats the initial "that goes" in quotation voice and tries to delete the characters from the VDS but finds that it is not possible to go "bahck uph" with the DELETE key. To resolve the problem, R suggests an ending "the cement works" that S

carries out. The possible trajectory of the incomplete construction is continually shaped by the syntactic and semantic material of the just prior contribution construction and the medium. The VDS record is a pervasive surface on which contributions must be constructed within the evolving circumstances of prior work on that surface.

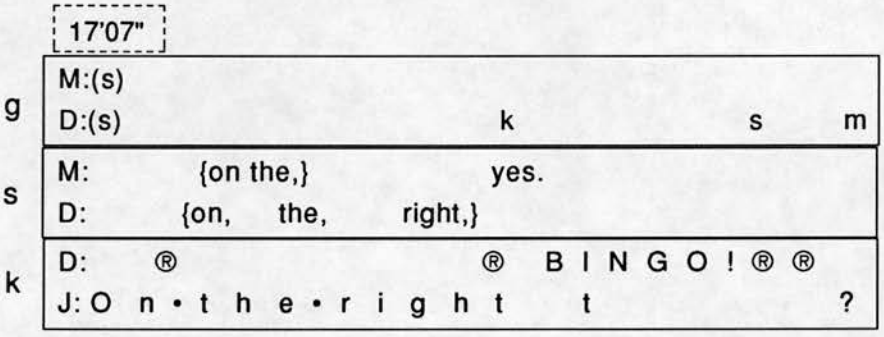
3. Occurrences of more than one speaker at a time are common, but brief (in conversation).

Participants do orient sometimes to a speaker, but commonly they are constructing at the same time. Constructions of contributions by both parties do occur at the same time and for quite a long period that are non-competitive, ie. they are not brief overlaps nor competing for mutual attention.

4. Transitions from one turn to a next with no gap and no overlap between them are common (in conversation).

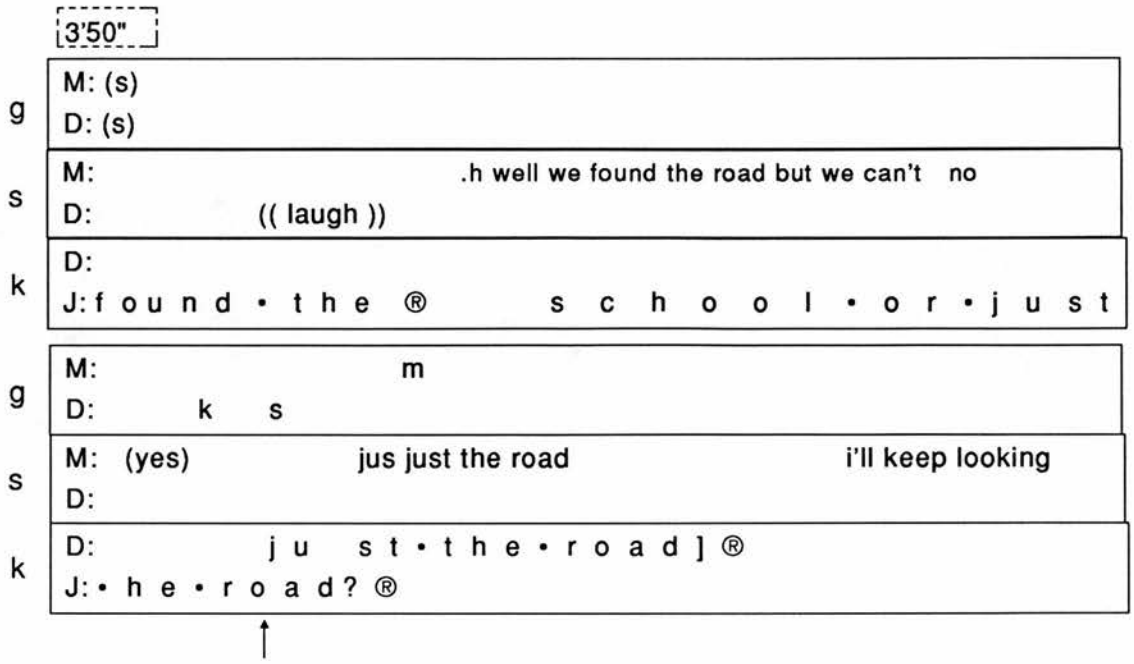
This is not true. Gaps are long when orienting to 'turns', from one to twenty seconds. Otherwise, constructions are regularly overlapped when the notion of a valued 'turn' of the floor is undermined. However, there are examples of closely monitored rapid or briefly overlapped transitions. In Example 22, the recipient's activity is closely coordinated with the other party's construction of a contribution. Both coparticipants are reading the current production by the other party. This means that they can coordinate their response in relation

EXAMPLE 22 - [M5: 17'02"-17'12"]



to the other party's contribution construction immediately they find an appropriate place to do so. In this case, a slight overlap occurs because of the projectability of "on the right" as complete enough for their purposes. The other party's construction confirms the success of locating a destination - M responds "yes" and D simultaneously replies with "BINGO!" in the VDS. Note that this is before the clear completion of the construction because it is misspelt and then transformed from a possible statement into a specific question after D's reply.

EXAMPLE 23 - [M9: 3'50"-4'05"]



Example 23 is similar to the previous but illustrates self-selection in the other party's construction that also collaboratively completes the construction simultaneously. M and D are closely monitoring the ongoing construction "youve found the" when M suggests a possible response projecting from the material so far. The contribution continues simultaneously with "school" and M responds "no" to her coparticipant. The item "school" now predicts and completes non-syntactically her own speech itself, eg. 'we found the road but we can't... school'. Note that the talk is embedded in the visual and temporal circumstances of looking at the display - her coparticipant could not understand the talk without interpreting it in sequential relation to the characters appearing and as a commentary on those characters as a contribution. The construction continues with "or just" and M predicts what the completing noun phrase will be but only indicates this by commenting "yes". Her coparticipant immediately looks to the keyboard and starts the completion with the same initial shape as the other party before the noun phrase is distinguishable. A successful completion is achieved which both completes the construction in progress with an identical and timely shape and also answers the question put by the other party in the process. This sort of self-selection is not always so smooth and will shall see that early self-selection is one reason why turn-taking similar to conversation becomes unstable and a new order of participation emerges in the VDS.

5. Turn order is not fixed, but varies (in conversation).

EXAMPLE 24 - [M5: 24"-1'20"]

```
→   D:   We have to guide ((D and M attend to other party's construction))
    J:          [W e    h a v e   a  starting point on a ®
          map of  cambridge, England ®
→   D:                  [you round certain ®
→         p
    D:    ok.
→         ((D deletes 'p' and scrolls window clean))
    J:    Do            y  o  u
→   D:       [We have the same map]
          ((D and M are now unsure of their statement))
    J:    have a similar map]
→   D:             [((D deletes 'the same map'))
    M:    oh. ((M has noticed the other party's contribution))
→   D:    ((D types 'a'))
    M:    {do you have a similar} (.) Put (.) Put (.)
→   D:                          [((D deletes 'We have a'))
    M:    we don't know (pause) probably
→   D:    We think so
```

This is the same in the VDS, but more regular than conversation because of the problem solving task. It has been observed that casual personal use of the VDS is more similar to conversational talk in its distribution of turns.

6. Turn size is not fixed, but varies (in conversation).

This is the same in the VDS because a particular phenomenon is the contingent construction of a turn. that is, a turn may be altered or abandoned or continued in a new trajectory because of the bursts of monitoring of the display for evidence of new actions, eg. a new chunk, or a contribution under construction, by the other party could lead to a revision of the current contribution.

EXAMPLE 25 - [M5: 12'30"-12'56"]

```
    Others:  Obviously you map is more detailed®
    D:       ®®BLAST
→   Others:  than ours
    D:       ®®
    D:    okay. ((D and M both look to map))
→   Others:  ((Others continue with 'however we have got to'))
    M:    <right, so they should be able to get onto coldhams,    lane.
    D:                                          [Can you see [coldhams lane.
          ((both look to screen and notice other party's construction))
          ((D types and deletes 'c'))
→         {we have got,}
    M:          [{however we have got to,}
```

Example 24 nicely illustrates the interactive construction of a contribution because of the monitoring of the other party's 'competing' construction. D starts almost simultaneously with the other party and types "We have to guide". D and M then attend to the other ongoing construction. After a while, in other party mid-construction, D continues on from the hung construction with "you round certain p". However, D abandons it in order to address the now complete turn by the other party. After constructing an appropriate response, "We have the same map", again in overlap with the other party, they have doubts and so some of it is pruned in order to continue. But, at this point M notices and reads the other's construction "Do you have a similar map". This fits in with D's revised proposed statement, but their next actions show that as the proposal stands it is not really addressing the question. Rather than continue with the incomplete edited turn, it is abandoned as suggested by M with "Put (.) we don't know", and an appropriate answer to the question is constructed. Therefore this example illustrates that contribution size is subject to local contingencies.

The empirical ambiguity of contribution completion points for the recipients is shown in Example 25. J responds to the turn construction, "Obviously you map is more detailed", with the exclamation "BLAST!". However, the other party had not finished the contribution because they continue with "than ours". The comment by J can then by re-interpreted not as a post-turn contribution but as a sort of backchannel to the main contribution in progress. Now, D and M both recognise "than ours" as a continuation that leads to a new completion point. D and M orient to the map, but the other party continues unnoticed with more, "however we have got to". At the same time, D and M decide on an appropriate contribution and D types the letter "c", the first letter of their proposed contribution. At this point, the other party's continuation is discovered, D abandons the proposed construction, and they wait for the final completion. The ambiguity of completion points is another reason for the slide from

EXAMPLE 26 - [M8: 4'02"-4'40"]

	G:	Directions to follow (pause) Left (pause) Blinco Grove ® ((both look to map))
	E:	o.k. ® ((K looks to screen))
	K:	okay. ((K looks to map and G to screen))
	K:	right um: ((G looks to map)) so it's (.) left ...
→	E:	how fsr? ((K and G do not notice))
→	G:	Then left...

turn-based dialogue into simultaneous and parallel contributions discussed in the section on 'double dialogues'.

7. Relative distribution of turns is not specified in advance (in conversation).

This is the same.

8. Number of parties can change (in conversation).

No, the parties are always two in number in the study, though this does not have to be so in any VDS. Multiparty computer-mediated communication is feasible though the synchronisation of multiple contributions in visual space is difficult.

9. Turn allocation techniques are obviously used (in conversation).

They are sometimes used. But Example 26 illustrates how the weak mutual monitoring characteristic of the VDS leads to unratified allocation. G first displays with a metacomment "Directions to follow" what is going to happen next and thus what they are engaged in doing and why there might be a delay to the other party. The route construction is then given as "Left Blinco Grove". After looking at the map K returns to the screen to find an affirmative response "o.k." by the other party. This is also recognised by G who looks to the screen. They then return to the map to construct the next direction sequence. Meanwhile the other party continues with a contribution "how fsr?". This goes unnoticed as G and K construct their next contribution. Thus both parties have self-allocated in an ambiguous post-completion position, but they are both unratified. Unratified allocation would be a constraint if the medium of communication was transient, because, in simple terms, coordination of production and reception is essential. However, the graphic properties of this modality eliminate this constraint. This means that communicative activity in the VDS tends to frequently engender simultaneous or 'overlapping' contributions that are unratified by the other party.

10. Various 'turn-constructional units' are employed (in conversation).

Constructions are not given as such but they have projectable characteristics as complete contributions. This leads to early self-selection and the loss of 'turn' orientation because of the unratified allocation.

11. Repair mechanisms for dealing with turn-taking errors and violations obviously are available for use (in conversation).

There are several of the typical repair mechanisms to be found in talk. For example, there are self-editing repairs, eg. repair of misspelling or turn misplacement by deleting or scrolling (using the DELETE or RETURN key). Also, there are some other-initiated repairs, eg.

expectable turn allocation take-up if a turn is missing in a 'reasonable' time ("go on !"), or echo-question repair of sequentially strange contributions ("ANG?"). But, many that are operative in repairing troubles in conversational talk are unavailable because errors or violations of turn-taking systematics transform the organisation of participation itself. For instance, self-selecting in overlap is a violation of the turn-taking systematics but in the VDS it is not resolved by the participants. It is demonstrably not problematic on many occasions and is thus a regular phenomena to be explained.

Let us return to some general observations about the VDS with respect to a systematics of

EXAMPLE 27 - [M8: 4'24"-5'50"]

```
      E:    o.k.
            ((K and G attend to their next construction and not to the screen))
→     E:    how fsr?=
      G:    =Then left A1307 mark Homerton Teachers Col
            ((K and G both look to the screen))
              (longpause)
→     G:    oh wait a minute (.) ok, how?
            ((K looks to map and then back to screen))
            what's this?
              (pause)
      G:    ( )
→     K:      [{how far.}
→     G:    how far! (.) right.=  ((G and then K look to map))
      K:    =oh.
              (pause)
→     K:    .hh if they've got- ((K looks to screen)) um-
            ((K looks to map, G to screen))
            ((K looks to screen))
      G:    let's see if they've picked this up.
              (longpause)
      K:    ((looks to map, back to screen and then to map again))
            to
Others: what
            ((G starts to read aloud, K looks to screen))
Others: road is it on?
K&D:  ((both return to the map))
```

turn-taking. In talk turn-taking there are two hypothesized components. First, a speaker is given a turn-constructional unit. Second, next turns are allocated in situ according to two principles: the current speaker selects next speaker or allocation by self-selection. The systematics is a description of structural resources that analysably inform the contingent accomplishment of participation in situ. But in the VDS the lack of a routine synchronisation of monitoring leads to the disintegration of the systematics and the 'turn' as an achievable phenomena. This causes problems for the coordination of sequential interpretations routinely

achieved in talk. But this does not result in the disintegration of the dialogue itself because there does emerge a different organisation of dialogue, namely 'double dialogues', that will be discussed later.

For example, an apparently completed construction can become a troublesome item when the recipient interprets it some time later because of delayed monitoring (see Example 27). After an apparently terminal "o.k." by the other party, K and G attend to the design of their next construction. Meanwhile the other party continue with "how fsr?". After they have typed their construction, "Then left A1307 mark Homerton Teachers Col", K and G both look to the screen. K notices first the unratified contribution and they both proceed to resolve this sequentially ambiguous item. For the other party it was produced adjacent to the prior direction instruction by K and G which was "Left Blinco Grove"; to resolve the extent of the distance to proceed along that road (not the A1307). The just completed response by K and G resolves this problem and so is an appropriate answer to their query. But, for K and G it poses a distinctive problem. Unwittingly it now appears sequentially out of turn and it is not clear what interpretation should be entertained. For example, it could be that the contribution is to be understood in relation to the immediately prior as they find it, which is "mark Homerton Teachers Col". Thus the item shows that the destination has not been found and clarification is required. Equally, the item could be placed as requesting elaboration of the distance to travel when following the direction instruction "then left A1307" before the completer "mark Homerton Teachers Col" was typed. But K finally appears to relegate it as an old item not relevant to their immediate route construction dialogue when she says "let's see if they've picked this up". But G is still visibly puzzled by the item. The other party respond with "what road is it on?" and this indirectly resolves their dilemma.

In writing letters, a turn at 'communicating' is oriented to by the participants in the correspondence itself, eg. a writer can indicate that a letter can be out of turn having crossed in the post, a letter can be noticeably absent or late, and a letter can refer to prior letters (Mulkay 1986). So, the concept of a 'turn' is found in the participants' displays and orientations to their own participation organised through the physical constraints of production, reception, monitoring, etc. The notion of 'turn' is a contingent property of the interaction, being worked on within multiple constraints. In general, participants cannot just act at any time in interaction, there must be physical or social resources for managing the sense of actions being in relation to others.

However, in the VDS the participants can save time and produce overlapping responses that weaken the analytic notion of 'turn'. Participant's tend to produce responses within their own time of monitoring - a response is due when a string has been interpreted as relevant contribution by the other party. However, coparticipants do orient to the appropriate time for

a response with respect to the other party, and thus adjacency is still a resource. The routine allocation techniques are weakened in the VDS and the shared floor reduces to parallel activities where parties can attend to a contribution after its projection or two streams of

EXAMPLE 28 - [M9: 3'42"-4'05"]

```
      J:        at the
              ((Success of earlier step becomes problematic in their talk))
                ((J deletes 'at the'))
→     Others: where
      J:        [youve found the®
      S:      (they're) typing something.
      J:        school or just he road?®
→     Others:                    [ju  st the road}®
→     S:      where just-
      J:      the road. (.) ok.
      S:      just the road, ok.
```

participation that are simultaneously interleaved. It is possible, as has been shown above, for a temporary return to an orientation to a 'floor' to occur.

So, participation in the VDS is a continual and contingent issue for the participants. They are borrowing the normal, everyday practices of conversational interaction in order to constitute the significance of and interpret actions. Participants find the string of characters appearing throughout the session sensical. To them, turns are constructed in front of their very eyes and become sedimented in the semi-permanent graphic window. Alternatively, new responses can be recovered from the graphic window after they have been produced by recognising changes that have not been seen before. The boundaries of actions and the continual classification of priors are locally constituted; maybe with multiple and changing interpretations. For example, in several data fragments there is evidence that prior or abandoned characters, in the process of constituting communicative actions, are interpreted as part of the current communicative action. Let us look at Example 28. Both parties are engaged in simultaneous turn construction. One party stops after "where", but J continues. A little later her coparticipant S notices the initial typing by the other party and comments "they're typing something". Towards the end of J's construction the other party collaboratively overlap and S initially reads their contribution as "where just". This is because in the VDS the abandoned construction and the next that is relevant at this point are on the same line with no punctuating marker. Thus the boundaries of communicative contributions is empirically problematic for participants in this case.

5.4.5 Contributions in a Dialogue: Conditional Relevance

In talk, turn-taking establishes a seriality to participation on which an adjacency assumption can be built. A current turn by one party is expectably oriented to in the next turn by the other party and that turn will be interpreted as such unless otherwise provided for. But in the VDS there is no mutual ratification of a 'turn' at typing and unratified allocation occurs frequently. The instability of weakened allocation leads to the disintegration of the singular seriality of dialogue, ie. there is no mutual floor and simultaneous construction is common. However, some sense must be found in the characters displayed in the relevant window on the display, or the absence of type that occurs, eg. it could be a machine error, or random typing by an unknowing participant, or a reply to a question, or an unwarranted interruption. The issue for the coparticipants is: What sense do the display changes have in terms of prior communication and the practical activity? It is perfectly possible for the participants to type continually or when they feel like it. This is also possible in talk, but does not happen. In order for shared understanding to be achieved and maintained in multi-participant communication and interaction, the relevance of actions to each other must be constituted in some way. My actions must relate to yours in ways that are mutually intelligible. In the VDS, the participants are continually achieving the sense of what an action is through its placement with respect to prior actions by themselves and the other party as they construct their participation.

EXAMPLE 29 - [M10: 11'14"-11'31"]

```
      S:    ...a jetty
              (longpause)
      R:    °ok°
              (pause)
            ((S looks to map))
      R:    .h  take the footpath: that goes past (    )
→     S:                                          [< i don't know if they've seen that
                                                    ((S and R look to screen))
            ((S looks to map)) yet=cause they havn't said anything.
→     R:    °right°.   we could type it ((S looks back to screen)) in anyway.
              (pause)
      S:    yes ok.  erm:  ((they both look to map))
      R:                 [um,
      R:    take the footpath ...
```

Example 29 shows that adjacency is a resource for interpreting the dialogue but rather than orient to an expectable adjacent response they continue on with further contributions. Participants S and R construct a direction terminating with "a jetty". R works on the map while S waits for a reply from the other party. Then R orients to a possible next activity, S

EXAMPLE 30 - [M5: 3'05"-3'30"]

```
D:      left or right side of map ?
          (longpause)
D:      ((looks to map))
→       o'come on (.) you've gotta be able to tell us that!
        ((looks to screen))
          (longpause)
Others: Can you see the Government...
```

The other party, J and R, are working on a possible response while D and M wait.

```
D:      left or right side of map?
R:      light- side of the \map
J:                          [right.
R:      no it's not  it's in the centre.
J:              [oh.]
J:      .hhhh
R:        [ah but half
J:      (hold on)
R:      between,
J:      if they've got...
          ...
→ J:    how can we describe it much better
          ...
J:        ((starts to type construction))
```

looks to the map, and thus mutually engaged, R launches into the next direction instruction. S overlaps towards the end, "i don't know if they've seen that yet", claiming that another contribution would be inappropriate at this point as no expectable sequentially adjacent response indicating receipt has been found in the VDS by S since the contribution was finished. They both inspect the screen for evidence and then R proposes that they can type it in anyway. S inspects the screen again and after a pause agrees "yes ok". They proceed to construct a follow up contribution.

In Example 30, the conditional relevance of contributions is displayed. D types the contribution "left or right side of the map ?" and they wait for an appropriate answer in second position to the question. After a while, D looks to the map and then exclaims "you've gotta be able to tell us that!" This displays, for D, that the delay signifies that the other party is having trouble answering, but not that they have misunderstood the first part of the pair. Finally, they respond with a question which can be interpreted in sequential position as building in some way on the second position slot, eg. that they aren't sure about the answer and require more information. However, as has been shown above, this cannot be relied upon given the weak mutual monitoring, eg. the other party may not have received the contribution and so their construction should not be interpreted in adjacent position. The analyst must inspect both parties activities to see how the contributions are to be interpreted

EXAMPLE 31 - [M8: 3'38"-3'55"]

```
Others: school located ®
E:      school located. right. so they've got that. mhm?
R:                                  [right]      [ok, °right.°
        ((E and R wait for a long time))
→ E:    (°and now what they gonna tell us.°)
→ R:    now they'll ask us another question presumably 'cause,
        (pause) they're the ones who have the instructions.
          (longpause)
→ R:    °or they'll tell us to do something maybe°
E:      mhm.
          (longpause)
E:      ((laugh))
          (longpause)
Others: Directions to follow
```

and what resources are required to do that. In this case, J and R have recognised the question and find they cannot answer either way - "it's in the centre". After some more talk J suggests they find an alternative that would "describe it much better" and thus places the work of constructing a contribution in the larger activity of finding landmarks in order to locate items on the map surface. Also, it is designed as a response to the just prior adjacent question and is interpreted as such.

An illustration of orientation to what is expected next is given in Example 31. After receiving a response that confirms the other party have got the starting point, E and R wait for a turn continuation - they do not self-select - for about 10 seconds. Then, E says softly, "and now what the- they gonna tell us". R responds with "now they'll ask us another question presumably 'cause, they're the ones who have the instructions". E and R have a firm conviction that it is the other party's turn, and do not initiate repair even when the period of inactivity grows long and noticeable. In the end, the other party, who have been working on a suitable route plan on which to base a response, do continue and thus resolve the initial doubts about what they are doing.

The above examples all illustrate that adjacency is still a strong resource for interpreting items or their absence in the VDS. However, regularly in the VDS contributions are not received, mislaid or misplaced and the relevance of an item to the evolving dialogue has to be judged much later, if at all. An example of a sequential misplacement trouble was mentioned in the last section with respect to the fragment [M8: 4'24"-5'50"]. In the next extract in Example 32, many contributions are no longer adjacent upon receipt, but which were originally placed in that temporally sequential position. Also, there are multiple conditional relevancies of which some are oriented to in talk and others in the VDS. This example requires an examination of the dialogue from the two different perspectives of the parties involved. The dialogue

EXAMPLE 32 - [M5: 5'55"-6'25"]

From the point of view of M and D.

J: Can you see Addensbrookes Hospital?
D: Not yet, we're working on it.
((D and M search for starting point))
D: We have a primary school at roughly the right place. How do you know what it's called ?
(pause)
Is it opposite baldock way ?
→ D: {are you blind?}
M: ((laugh))
D: are you stupid!=
→ M: ={can you see the station?}
D: can we see the station=No th- we've got the primary school.
...
J: Good, yes
D: Right, we have to guide...

From the point of view of J and R.

J: Can you see Addensbrookes Hospital?
D: Not yet, we're working on it.
J: [Are you blind?
J: Can you see the station?
D: We have a primary school at roughly the right place.How do you know what it is called ?
→ R: oh the name's on our map.
D: Is it opposite baldock way ?
→ R: yeahs, it's here.
J: Good, yes correct
D: Right, we have to guide...

marked on the examples represents the temporal order of receipt. First, from the perspective of M and D, the other party asks a question to which they respond in adjacent next position with "Not yet, we're working on it". They then search for the starting point on the map surface, find it and respond with one statement and two questions in quick succession. Each of these has some expectable response in next adjacent position. For example, the first could be followed by an assessment or confirmation, the second by a full answer, and the third with a yes or no. While waiting D finds a contribution "are you blind?" by the other party on the display. This is treated jokingly and not responded to in the VDS. M then finds another contribution "can you see the station?" which is also not responded to. Neither of these anomalous contributions is routinely interpretable in adjacent position because of the lack of mutual monitoring. Neither have they been designed with their misplacement in mind as it is an emergent and unwitting property of the interaction. Finally, the other party respond, in

the face of three possible nexts from D and M's point of view. The response "Good, yes" can be interpreted as orienting to the first and third according to their expectable formats. D then moves onto a next activity in the task.

From the alternative point of view of the other party, who are J and R, J asks the question: "Can you see Addensbrookes Hospital?", and the others respond negatively. In the context of the dialogue so far they have not been very successful so J asks jokingly: "Are you blind?". To help the search, J then asks another question: "Can you see the station?" This treats the previous contribution as minor and now superseded. D then apparently replies that they may have found the starting point and asks a question indicating a problem with the name. It is ambiguous as a reply because it is not clear in the VDS whether adjacent positioning is a valid resource. Thus, it cannot be determined whether or not the question about the station helped them find the starting point. In this case, it did not and is only found later in the other party's activities. The question is answered verbally by R, and the next question is also verbally confirmed. J then responds "Good, yes correct" simultaneously dealing with the first and third statements and generally tackling the wave of sequentially implicative contributions that may be operative.

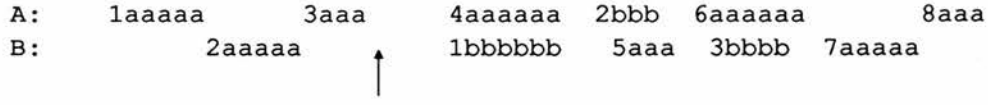
5.4.6 Emergent Organisations: Double Dialogue

The phenomenon of 'double dialogue' is an interesting dynamic in this modality because of its properties and the social resources methodically deployed in achieving sense. Basically, a two threaded dialogue is achieved as follows. Normally, participants in speech conversation achieve an accountable sequential adjacency through the operation of a 'turn-taking system' by which one person speaks at a time. Sequential adjacency is still important in the VDS, however, parties do not necessarily need to monitor transient events and so more than one person can 'speak' at the same time. Either party can produce a contribution as a response to some contribution by the other party, while the other is doing the same but to a different topic, because the contributions can be produced as complete and 'to be read a little later'. Topics can be developed simultaneously and can intertwine in a way that is not feasible in spoken or signed dialogue. Simple examples occur in the data because of the type of task they are engaged in, but conversational uses of the modality have been observed in which extended 'double dialogues' are maintained. A detailed illustration of the maintenance of extended 'double dialogues' is not dealt with in this thesis. Examples DD1 and DD2 will illustrate. Two topical threads, indicated by the letters 'a' and 'b' or 'c' and 'd', are maintained simultaneously. The numbering is incremental by turn within each thread and time progresses from left to right. At least two types of onset are possible and are found in the early stages of double dialogues in the corpus.

One type of onset happens when there is an ambiguous completion as shown in Example DD1. This could occur when A types '3aaa' for different reasons. After a pause, A continues at the same time as B, but neither knows about the others onset and so they are not turn competitive. This has been shown to occur in the analyses above, eg. Examples 14 and 26. At the termination of their respective constructions, each attends to the other's completed contribution and responds to it at the same time. And so it continues until one attends to the in progress construction of the other rather than respond to the prior.

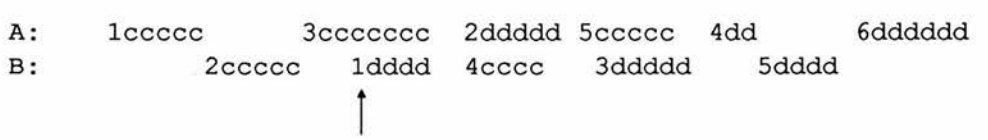
In Example DD2, a second type of onset is illustrated. During A's '3cccccc' construction, B self-selects early and starts a contribution that is predicated on a certain completion trajectory of the interrupted construction. This turns out to be incorrect and thus two topics are now potentially operative. This has been documented in the analyses above, eg. in Examples 14 and 24. Upon completion, each party can address the other's topic simultaneously, as before. A good example of the ambiguity of a contribution completion point that leads into the beginnings of a double dialogue is given in Example 33. D asks a meta question about the requirements of the task, but the other party indicates there is a problem: "Stop". They elaborate that they are lost. D then recycles the meta question thus showing that it was not answered and is still relevant, and also claiming that it is valid for them to raise such an issue at this time, eg. they have a right to a contribution. However, the other party also claim a right because they construct a further elaboration of their problem. Thus the contribution "we are lost" is problematic in terms of providing material for each party to coordinate their next

EXAMPLE DD1 - Possible double dialogue onset: ambiguous completion



actions. But the medium allows simultaneous typing and so both parties construct contributions dealing with distinct topics. Immediately after completing the recycle, M finds the other party's construction relevant. At this point the double dialogue unwinds prematurely but if D and M had constructed a contribution in reply then it would have overlapped with the next response by the others which answers the recycled meta question, with "Of course". Thus two topics would be in simultaneous and alternate progress in the VDS. Instead D attends to the screen and waits for a reply to the recycle which is forthcoming. Only then do

EXAMPLE DD2 - Possible double dialogue onset: early self selection



they construct a contribution addressing the problematic directions and missing cement works.

The occurrence of double dialogues is not surprising given that parallel or simultaneous activities by the parties engaged in the VDS are prevalent, eg. as documented in Examples

EXAMPLE 33 - [M5: 9'55"-10'25"]

D and M have just given a route direction.

```
D:        Are you marking the route too?®
          ((D looks to screen))
Others:   Stop,
D:      {STOhP!} ((M looks to screen))
Others:   we are lost
D:      {we are lost:}
M:            [loszt:
M:      ah:::
D:       [oh.
Others:   ®
D:        ®®
→ Others: W e   h a v e   n o   c e m e n t   w o r k s
→ D:         [Are you marking the route too?®
M:      {we have no cement works?}
        ((M and D look to map))
        ((D looks back to screen))
Others:   ®Of course
D:      {of course.} ok. ((M looks to screen))
D:        ®OK. From the ...
```

21 and 27. If each party can attend to divergent activities simultaneously, eg. both are looking at their map at the same time, and if one party can attend to another activity, eg. looking at the map, while one party constructs a contribution in the dialogue, then it could be possible that both parties attend to their own dialogue constructions while disattending the other. That is, one party can engage in the activity of constructing a contribution at one and the same time as the other party are also engaging in constructing a contribution to a different topic but to the same dialogue. Note, there is a difference between parallel and simultaneous dialogues: parallel dialogues are simultaneous but are concerned with the same issue or problem, so double dialogues are parallel. Another sort of parallel dialogue organisation called 'byplay' occurs commonly in multi-party gatherings when hecklers conduct simultaneous commentaries with other recipients on the ongoing 'legal' talk that occupies the floor (Goodwin, personal communication). Other forms of simultaneous dialogue activity are emerging in the research literature but none have the property of focused intertwining of topics simultaneously by only two parties.

Unfortunately, the data in this study was not very good at showing extended double dialogues[9]. A fast dialogue situation would encourage more highly developed examples. There are many in conversational use of the modality and they were first observed in pilot

EXAMPLE DD3 - Digressions in conversational talk

```
A:    1eeeee      3eee       4eeee      2fffff      6eeeee
B:          2eeeee      1ffff      5eeeee      3fffff
```

studies and observations of conversational uses of the VDS by competent routine users. In these cases, sequential adjacency is still a very important resource in achieving mutual intelligibility. However, it is not a mutually monitored temporal adjacency as in spoken conversation in contexts of copresence but a reconstructed adjacency from the perspective of each participant that emerges in the dialogue itself. The participants contingently achieve two threads of dialogue each with their own sequential development.

Is 'double dialogue' possible in speech? No, because of the real time coordination of speech production and reception. Probably, the nearest example is for both participants to be

EXAMPLE DD4 - 'Theatre of the absurd' play dialogues

```
A:    1ggggg      2ggggg      3gggg      4g      5ggggg
B:          1hhhhh      2hhhhh      3hhhh 4hhhh
```

alternating topics, but only one participant will be in phase with the other (see Example DD3). This is quite a regular occurrence in copresent circumstances where several activities are possible, one verbal and one physical. One activity may be treated as a digression or side sequence. Occasionally participants do overlap a number of speech segments as they complete their turn construction and allow the other party brief turns at talk. But these are competitive non-attentive moments in problematic talk. Double dialogues are not competitive nor problematic practices. Is 'double dialogue' possible in fictional dramatic dialogue plays? You can have patterns where the participants ignore each other's content but appear to interact by managing the turns in a normal conversational manner. In Example DD4 above both A and B are not visibly paying attention to each other's topic, though they may be, but are still managing to coordinate turns at not paying attention to what each other says.

9 It is not clear from the corpus whether or not the dynamics they were engaged in was a necessary step towards competent double dialogue use.

5.5 Summary and Relevance to Computer Models

This chapter has presented an analysis of the achievement of communication in the virtual dialogue space, which is indeed a skilled contingent accomplishment by the participants wrested from the circumstances and emergent dynamics of their mutual engagement. For example, without routine mutual coordination the sequential placement of contributions becomes problematic. Crucial features of the VDS are: the visual, temporal and spatial qualities of the VDS in contrast to spoken talk; the permanence-transience qualities of the medium; the dialogue distance between participants that allows them to treat the dialogue not only as a mutual focus of attention but also at other times as a disengaged distributed activity; and the emergence of new orders of participation. The VDS is a problematic yet skilfully coordinated achievement that shows it is not possible to predict what organisations will emerge in odd or particular circumstances.

Dialogue in the VDS routinely has quite a different organisation though it can only be understood with reference to the organisation of conversational talk. The VDS has certain characteristics which are in flux because of the weak mutual monitoring and permanence characteristic of the medium, and thus the activity is not quite like talk or writing. However, the dialogue does border on being similar to writing correspondence at times, and at these points the reduced characteristics of interchange dialogue are evident. In the VDS the constraints characteristic of interchange dialogue are lost. No longer is there a strict turn boundary delimitation, nor a single focus of mutual dialogue activity. Because of the flux between different participation orders, the complexity and crucial features of achieving that participation are revealed. The analyses illustrate three gulfs or simplifications inherent in the interchange model of dialogue.

1. (a) 'here is what you have been waiting for' - like wrapping up a message in private and then handing it over. The in situ chronic work of producing a message is not available as a performance. Only the residue or product can be inspected and it is so designed by the participants. This is not an inherent characteristic of the VDS, but can occur when participants are engaged in parallel activities.

1. (b) 'take it or leave it' - like wrapping up a message in public, but without recipient design. This is a less extreme form of (a). A contribution can be performed with a recipient audience, but design of that performance cannot be altered on the basis of attentiveness to the recipient's complementary actions. This is a contingent characteristic of the VDS, but the trajectory of ongoing contributions are regularly altered in response to the other party's constructions.

2. 'wait until I say' - like exchanging messages with the rule, "give me a

message only when I have given you one". The current turn-holder is in full control of turn allocation which can only be allocated on the explicit signal of that person. This is not the case in the VDS because participants must locally construct their participation.

In the restricted and non-everyday modality encountered, there is a rapid adaption and development of resources and organisations. New circumstances and contingent emergent problems lead to a tottering achievement of sense through borrowings and the documentary method. Rather than the routine and transparent flow of conduct, special interactional jobs need to be done. Some are recurrent and lead to sedimented forms of conduct and practices. This finding supports Wynn's claims that there is "a locally meaningful and workable background of practices that members of communities continually re-negotiate and re-establish implicitly in the course of performing" (1979, p. 7) and that "the discourse supporting activities of natural conversation always address practical concerns." (1980, p. 89)

This chapter has shown that an empirical methodology can fruitfully examine the situated practices of those involved in dialogue. With respect to computer models of dialogue, the analyses have shown that interactivity is a constitutive domain of dialogue that cannot a priori be ignored. Orders of interaction should not be dismissed from a theory nor taken for granted, otherwise the simplifications pin-pointed above become the properties of dialogue implicitly reproduced in computer models, like that of Grosz & Sidner (1986), and interesting phenomena are thus excluded. For example, the situated achievement of dialogic presence and monitoring, as well as the local management of participation in presence from the materials at hand, have been taken for granted. Thus aspects of the performance of action, the emergence of participation orders and the texture of interaction are ignored in the model so that a set of constraints on interpretation are lost. In addition the data used in the empirical justification of many models has not only lost a sense of interactive life, but also of temporal life. That is, an authentic dialogue may have characteristics of the interchange model, but the model renders it in abstract time. The danger is that the interchange model and its underlying ideology are to be reified in the actual design and conception of dialogue systems based on such models. The wealth of interactional resources that give priority to the visibility and local achievement of action could be useful for systems which provide limited access to already depleted circumstances and which over-represent and over-rationalise resulting in unwieldy inferential mechanisms.

The next chapter will look at some of the interactional resources that may help the creation of interesting novel artifacts. In addition, the methods used in the study are also applicable to the evaluation and analysis of people's use of such artifacts. The first step required is a

naturalistic analysis of routine human collaborative conduct with artifacts in a workspace. The chapter will take a careful look at the coordination and intelligibility of multiple human activities in contexts of copresence that interleave talk, body and artifacts. In combination with the analyses explicated in this chapter, which are restricted in the manner of possible artifact modalities a perspective on computer dialogue artifacts is presented. Providing for interaction will yield a wealth of resources and positive constraints.

Chapter 6

DESIGNING FOR INTERACTIVITY WITH DIALOGUE ARTIFACTS

6.1 Introduction

The study reported in Chapters 4 and 5 has demonstrated that an empirical methodology derived from conversation analysis (CA) can reveal the situated interactional work of constituting dialogue 'through' machines. This adds to an understanding of how people manage to coordinate and communicate in circumstances quite different from everyday talk. In combination with Chapter 3 it points to a range of interactional phenomena that have been taken for granted or ignored by computer models of dialogue. Also, the study exposes a number of features that can help in the design of computer-mediated communication systems. The modality is, however, quite similar to what could be expected in near future computer system design which supports dialogue 'with' machines. What consequences do the arguments of Chapter 3 and the study have for dialogue 'with' machines?

It is argued that an unnecessary restriction is placed on the design of systems supporting dialogue 'with' machines if interactivity is ignored. Contemporary computer models of dialogue contribute to the weak theoretical conception of just what dialogue is and how machines and humans could engage in dialogue. Insight into human dialogue practices demands that the interchange model be abandoned and the resources and dynamics of interaction recovered. Given the reconceptualisation of dialogue argued for above, it is necessary to examine carefully the situated and circumstantial coordination of action and achievement of mutual intelligibility rather than reproduce specific human activities such as everyday natural language conversation. It must be remembered that users ground and sustain the metaphor of dialogue 'with' a machine and thus a shift to a human-centred perspective that supports user's work and sense-making practices in their activities conducted in and through the machine would be more appropriate. The activities of users are also available for study, and the empirical methodology used in the investigation, eg. the transcription methods, could be applied to evaluating and analysing dialogue artifact use.

It is a claim of this thesis that designing for dialogic presence and interactivity will be one major contributor to more appropriate dialogue artifacts. A priority must be to eliminate the restrictions on dialogue interaction that come with the interchange model. But this does not just mean that the user or system be given opportunities to influence the construction of an activity, eg. the monolithic textual explanation found in manuals should be broken down in computer help or advice systems and the user given opportunities to influence the presentation of an explanation at the time it is required according to the user's needs. In addition, it is argued in this thesis that the construction of dialogue contributions or participation itself must be locally managed and subject to interactional pressures.

In order to gain some insight into what might be possible and the problems to be faced, a fragment illustrating the complexity of human copresent interaction will be briefly analysed. In Chapter 5 the analyses documented how participants contingently achieved dialogic presence and coordinated monitoring, and thus how performance features and a texture of interaction became important for constructing and interpreting action. How their participation was locally organised out of the materials at hand resulting in a shift towards parallel activities and emergent orders of participation, such as double dialogues, was discussed. Two more features to those found in the empirical study are isolated here: the complementarity of actions, and the embodied achievement of performance. The first is additional to reciprocal action and adjacent interpretation and is found especially in sign language conversation and copresent talk where actions by the same or more than one party are simultaneously produced and display an orientation to the other's actions. The second is the perceptual locus of the production of action connected with the body that is lost in the production of cultural artifacts. For instance, the monitorable shift into and out of action that provide resources for coordination of recipiency and action by other parties, eg. the body, gaze and arm movements before the acme of a gesture.

From the empirical study and its findings a number of recommendations will be made. First, because of the nature of dialogue artifacts, the sustaining of dialogue 'with' machines and the emergent properties of dialogue in more interactive circumstances, the user's behaviour must be channelled. The system's poor access to the resources of action can be alleviated by guiding the user in predesigned ways. Second, interactivity should be a basic consideration in the conception and design of dialogue artifacts. Artifacts should support dialogic presence - a background of mutually accessible 'visible' communicative resources and circumstances - and the performance of action which is a monitorable and observable embodied accomplishment with emergent characteristics. These features will provide an experiential texture for the coordination of simultaneous and parallel actions and activities.

The recommendations of this chapter pose a challenge for design that could be incorporated into traditional models, but a shift towards embodied systems active in real temporal-spatial contexts is necessary. However, it is hard to make generalisations about how to design computer environments supporting the metaphor of dialogue with a user because in human conduct dialogic interaction is a contingent accomplishment on each occasion. Any alteration in the normal circumstances and thus the resources of dialogue activity will result in the unintuitive disintegration and transformation of that activity as the study reported in Chapter 5 has shown can occur. In the VDS the notion of 'turn' and a mutual floor of speaking in the construction of participation lost sense as participants conducted parallel activities and moved towards a different order, viz. double dialogues. So, certain characteristic features of human dialogue will be not be reproduced in the resulting dialogue 'with' the machine. It is more than likely that in trying to achieve a semblance of dialogue presence by design, the dynamics of the emergent local activity will neither be predictable nor stable. The feasibility or appropriateness of these suggestions is of course an empirical issue that requires extensive research, design and evaluation.

6.2 Consequences for Dialogue Artifacts

> *Practically, ingenious design combined with testing may do much to extend the limits of useful machine behavior. Theoretically, understanding the limits of machine behavior challenges our understanding of the resources of human action. Just as the project of building intelligent artifacts has been enlisted in the service of a theory of mind, the attempt to build interactive artifacts, taken seriously, could contribute much to an account of situated human action and shared understanding.* (Suchman 1987, p. 189)

6.2.1 Conception

It has been shown that certain aspects of computational modelling of dialogue are problematic, and that the adaption of conversation analytic methods to form an interactional analytic methodology is suitable for the analysis of a virtual dialogue space. To return to the issues raised in Chapter 3, what consequences do the last two chapters that analyse the achievement of a VDS have for dialogue artifact conception, design and evaluation? A major aim of Chapters 4 and 5 was to illustrate that dialogue interaction is more than just the repeated exchange of dialogue objects between abstract entities, as dialogue artifacts have tended to be conceived as supporting. Models are used as a theoretical platform on which to design and conceive of practical systems that are intended to engage with a user in 'dialogue', and therefore the ideology can be instantiated in those systems. For example, the following definitions by Edwards & Mason (1988) are indicative of the weak conceptions of dialogue prevalent up to now in human-computer interaction (HCI) research. Dialogue is defined as the "temporally limited interaction between two or more relatively autonomous

entities through a sequence of exchanged messages" (p. 140). Messages "can be thought of as a collection of data of arbitrary length which passes uninterrupted from one dialogue partner to another" (ibid).

However, the unity of a 'message' and the boundaries between them that are very clear for the participants (and observer) in interchange dialogue systems becomes less precise in dialogic circumstances approaching copresence. The 'interchange model', implicit in the conception of dialogue that is pervasively present in computational models, ignores the importance of interaction for the everyday achievement of intelligibility in dialogue. There are two very good reasons for not using the interchange model. It produces dull artifacts with strict action sequences requiring all dialogic work to be done within the explicit speaker-only turns that make up the sequences. Also, there is an unavoidable deficit of important constraints on interpretation. For example, the absence of activity in a locally organised interaction can have a significance not possible in temporally abstract interchange dialogue. This has been documented in conversation where different types of significance can be attributed by the routine operation of the turn-taking systematics to an absence of speech activity, eg. as a noticeable absence of a response, or a gap before a speaker continues, or a lapse of turn-taking itself (Sacks et al 1978).

Edwards & Mason do acknowledge that "as human-computer dialogue approaches more closely that of human-human dialogue, characterised, for example, by pauses, interruptions, simultaneous speech and gesture, etc. the definition of message... will have to be re-examined" (1988, p. 144). However, this raises two related issues. First, is human natural language conversation to be taken as the model for conceiving of dialogue artifacts? Hayes & Reddy (1983) claim that a more artificial style of communication is appropriate because "sticking to human conversational conventions does not allow all the possibilities for communication between a computer and its user to be exploited" (Hayes 1983, p. 229). But they argue that this objection will disappear when speech systems can be integrated with other modalities. I would like to argue that both of these opinions support a rather restricting view of dialogue that elevates human spoken conversation as the ideal, superior form which systems must ultimately be competent in as conversational partners. It is true that human conduct and conversation is one very useful source for the creative design of some artifacts that may have similar characteristics. Also, users may draw upon the resources of talk in order to interpret and use the system. But the development of dialogue is a locally occasioned accomplishment subject to practical contingencies which result in emergent orders of interaction and dialogue.

> *The discourse supporting activities of natural conversation always address practical concerns. If a new concern should arise because of new constraints*

— eg. that the interactant is a machine — these will be incorporated into the ongoing details of communication. (Wynn 1980, p. 89)

Consequently, the simulation of human conversational behavior is not preferentially supported here. The primary concern instead is with the nature of dialogue practice by which mutual orientation and intelligibility are achieved and sustained in situ out of the materials available. Besides everyday spoken conversation, there are other surprisingly different human dialogic activities that can provide models for design that do not have to based on assumptions of face-to-face interaction, a sequential medium, or speech specifically, eg. sign language talk, electronic mail, back-to-back talk when occupied with other activities, deaf-blind haptic languages, and Bliss symbols for the physically disabled. The issue is not whether it is appropriate to build 'conversational' computers, but what is involved in the successful welding of computational power and modalities with human interpretative powers. A basic aim is the recovery of a background of presence in which interactive dialogue activity is constituted. Also, unfocused - parallel or simultaneous activities - as well as focused interaction will be common. Dialogue will take place in a virtual world of other entities and artifacts with a variety of modalities at hand. An understanding of how people coordinate their participation in similar contexts will help greatly in conceiving of appropriate dialogue artifacts and environments, and a preliminary study is undertaken later in this chapter.

The first issue implicitly raises the second: is the artifact best conceived of as an active partner whose competence approaches that of people? Because of the concern with conversational versus artificial dialogue artifacts, arguments have tended to assume that much machine use is dialogic. Care must be taken to make clear what should be considered as dialogue[1] activity and what is simply the active engagement by the user in activities mediated or supported by the machine interface. For example it was shown in Chapter 5 that Reichman (1986) had tried to explain a user's problem with an interface in terms of the conversational metaphor. Instead a simple explanation was offered on the basis of the fact that the user is engaged in solitary activity mediated by the machine as artifact, ie. the problem for the user was best understood in the context of the user's activities in and through the artifact than as a communication failure in a dialogue between the machine and the user.

1 The emphasis is on dialogue 'with' a machine as the dialogue activities that can be supported by the machine will be quite different in structure and content from dialogue with people.

6.2.2 Presence and Interactivity

A conception of the dialogue artifact is now possible that concentrates on situated interactivity. Chapter 3 argued that the design of dialogue artifacts can be conceived of as attempting to recover the practices of the local and situated determination of meaning. Texts are distanced cultural objects by which meaning is constituted in the play of differences located in the text itself. A competent reader will find sense in the linear text product in a socio-historical context. With dialogue artifacts the designer seeks to produce intelligible language in situ drawing upon the model of presence. Thus, in order to understand a particular piece of language that appears in the course of system use it is essential that its sequential placement or its situated performance in relation to the actions of the user be recognised[2]. Thus an understanding of how mutual orientation and presence is achieved, managed and sustained in human conduct will be of great help. The analyses in this thesis contribute to such an understanding. The phenomena investigated in Chapter 5, eg. the achievement and coordination of dialogue presence, monitoring and participation, illustrate the situated and circumstantial resources drawn upon by users, that may guide the design of interactive dialogue artifacts and that are missed in computational models.

Dialogue artifacts should support dialogic presence, eg. a rich context of interaction giving a background of 'visible' features that allow the coordination of monitoring in dialogue activity, and interactivity, eg. opportunities and possibilities for organising the mutual participation in dialogue. Studying interactivity gives a chance for routine and monitored processes that are situated in the flow of conduct to be incorporated, that hopefully deal with the issues relevant for practice and not the theory-specific problems of mainstream computer models of dialogue offer, eg. isolated cognitive processors desperately trying to communicate by recognising the intentions of the other participant as part of an internal 'understanding' procedure. The cognitive process view must recognise the situated, interactional and circumstantial nature of cognition and action. Computational models of language in use, eg. dialogue structure and cognitive process, are missing an account and recognition of the importance of interactivity and the situated flow of conduct. However, it is not claimed that the interactional perspective here is to replace the cognitive models which might be valuable abstractions for other purposes.

2 In some sense the placement can be interpreted by the user as an interpretation of the user's prior actions.

One consequence of the non-interactional perspective is that many phenomena are explained using models that must operate on much reduced evidence because the circumstantial production of action has been ignored in the application of the interchange model. Instead there are many reexplanations that can rely on external circumstances and are non-individualistic accounts, ie. they are emergent dynamics of the interaction, rather than the over-rationalisation and time-consuming computational processing characteristic of dialogue systems that are based on the interchange model. Consider a number of bizzare scenarios in which the resources of action that are taken for granted are made problematic. Imagine trying to shake hands with a stranger if only discrete moments of monitoring and action are possible, ie. hands can only be observed at specific moments without monitoring of the movement in the interim periods. Imagine trying to coordinate a conversational stroll in a park with someone if only the alternation of participation is allowed, ie. only one walker at a time can walk or adjust the pace. Imagine playing a ball sport, for example table tennis or tennis, where your opponent is not visible. It may be constituted as an alternation of discrete single moment opportunities for action in the game, but awareness of your opponent - body position, movement, balance - at all times is essential in order to play a skillful game with anticipation. Dialogue artifacts based on the interchange model are built so as to play ping-pong with an imaginary partner - utterances arrive and are produced as discrete units. Thus the designer has to construct a system that can recognise, plan and infer the user's actions and intents on the basis of severely limited evidence.

There is much 'out there' in the situated interaction with others and the world in which the sense of action and 'knowing how to go on' is found, eg. the familiar sense of finding things in situ but not being able to describe them beforehand, or, in conversation, not knowing what to say next but finding a next in situ, or not being able to be talk about when exactly to say something nor when to look at someone but doing it transparently. It might be argued that the tacit and contingent nature of much of our dialogic activities is unsuitable for computer implementation, representation and control. But this is not so if one thinks of the contingency of situated action not as a hinderance or problem to be avoided or overcome by design but a source of the richness of action and meaning. It would be a dull world to play in if the flux of the world was really a world of states and pre-constituted objects as mainstream AI would approach it. A characteristic of performed actions in interaction is the emergence of events as resources in the very act of doing, eg. the pace and rhythm of conversational conduct. So, the recommendation is to avoid individual representation and 'out of the moment' planning of discrete steps with only abstract sequential adjacency as a resource for interpretation.

Similar notions of interactivity are developing in research on interactive media technologies, like interactive video. The advent of direct manipulation technologies has shifted interface

design to the investigation of virtual dynamic media or spaces in which users have an active role (Laurel 1986). The traditional conception of an 'interactive system' is a reactive environment based on the interchange model in which the user formulates a command and the system reacts. By decreasing the grain size - the smallest element below which complementary or interruptive participation is not possible - an 'open' world of opportunity for action in the environment can be partially created. An appearance of unconstrained situated participation is sustained by having small grain moments of opportunity to act and many choices of action in each moment. Thus the user feels that she can actively participate at all appropriate times. For example, moving icons over a virtual desktop surface with a mouse, or driving a car around a town exploring side streets and roadside houses as the user wishes in an interactive video simulation. Simply, a user of a good system feels that direct, continuous engagement is possible[3]. However, if the system has too large a grain size, then only an interchange of explicit chunks is allowed, within which interruptive participation is not possible. If the choice is narrow, then it is not a very interesting world.

It is not quite correct to say that systems have avoided some of the features recommended in this chapter because they do unavoidably have peculiar and distinctive characteristics that users orient to and draw upon in routine use. But these are not necessarily design features. For example, the sound the computer makes or the length of time for certain operations in daily experiential use are background features from which odd noises or lengthy delays reveal problems to be addressed. Also, systems can be interrupted without negotiation using special commands from the keyboard or reboot, and many systems do not strictly adhere to the 'wait until I say' principle of interchange dialogue. For example, most systems have a means of buffering input and output, which to the user gives the impression of opportunities for action during the normal cycle of a process. Because of the separation of central processing from input/output devices the user can continue to construct commands. However, all input is held in a buffer and is operated on in the state that is current and not the state operative when the user constructed it. Thus the grain size is still the same and the user cannot interrupt the computer's central processing even though keyboard entry or other input device activity is possible. The user's input and the system output are buffered, creating

3 This phenomenon is analogous to the way films and video give the appearance of a continuous movement, by exploiting the propensity of the human vision system to see a stable, continuous image and not a flicker of distinct frames. If the speed of projection of sequential frames is fast enough (about 16 frames a second for film movies) then it works but if the speed is lower then the illusion is lost, ie. the individual frames become visible and are no longer coherent as a moving image. For interactive technology the notion of a threshold above which continuity is perceived applies at the point when a user of that technology feels that they can take the appropriate action without any apparent constraints.

a temporal distance between observable effect and production that can cause serious troubles and user impatience.

6.2.3 Dialogue Artifacts and Interaction Analysis

The study conducted as a major part of the thesis contributes to a developing interactional perspective on the situated achievement of mutual intelligibility. Because the dialogue was mediated through computers and took place in similar circumstance to that envisaged for a possible dialogue system it is argued that the methods and findings of interactional analysis are applicable to the design and evaluation of dialogue artifacts. In particular it is claimed that the design of interactive artifacts is attempting to represent and thus re-instantiate a vocabulary for talking about and interpreting dialogue artifact behavior and thus will have to do the same for the findings of an interaction analysis derived from conversation analysis (CA). See McIlvenny (1990) for a comprehensive discussion of this idea.

The notion of re-instantiating is meant to suggest the following idea. Conversation analysis studies the micro-details of conversational interaction, ie. looking for regularities that are demonstrable in the data, and positing simple, locally managed practices whereby conduct is made accountable and sensical. It takes traces of embodied and situated conduct that has occurred naturally and subjects them to a data-driven analysis. The aim is to uncover organisations demonstrably oriented to and reproduced in the conduct itself, ie. to describe and in some ways explain the dynamics of interaction in terms of the work done by the participants. Research in empirical investigations of communicative action disembody and objectify the phenomenon - to bracket it, talk about it, and understand it. Therefore CA does not specify how an action is to be carried out nor how an actor should 'go on' in a particular instance. This is not a fault of CA because it explicitly does not aim to explain the motivations or mechanisms that may generate the conduct under investigation. Rather it investigates how people construct and account for their behaviours as intelligible courses of action.

Because a computer dialogue system must perform, procedural mechanisms for generating its own behaviour and processing the behaviour of a user must be made explicit and be computationally feasible. What dialogue artifact design can do is represent the findings of a descriptive enterprise as structures for constraining the production of appropriate behaviours and for modelling the user's actions. Thus an attempt can be made to generate and 'give life to' the findings of CA as relevancies and phenomena reproducible in the user's dialogue 'with' the machine. For example, the turn-taking systematics is a context-free mechanism locally managed in context-sensitive ways to organise mutual participation. The description of the phenomenon in the ocnduct is accounted for by the systematics on the condition that it describes methods or resources that participants use to locally construct their occasioned

conduct. Dialogue artifact design can draw upon the systematics as a representable structure that is used procedurally to generate and thus hopefully re-instantiate the phenomenon it describes, ie. that the machine will engage in turn-taking behaviour and maybe orient to the 'turn' as people do. It is not at all clear that such an enterprise will succeed in the context of dialogue 'with' machines and in the concluding chapter an alternative that considers re-embodying foundational aspects of interaction and communicative activity in dialogue 'between' machines will be briefly presented.

The idea of using the findings and methods of CA as the basis for designing in HCI and AI research was discussed in McIlvenny (1985) in the context of a prototype study of advice-giving dialogue, and by McTear (1985) in the context of the repair of breakdowns in human-computer interaction. Demonstration of the suitability of conversation analytic methods for analysing the use of interactive artifacts has been undertaken by Suchman (1987). For Suchman the question is: What resources do users draw upon in interpreting their interactions with an expert system interface, and what trouble does this engender? She concludes that users of the system investigated do draw upon some of the communicative resources found in conversational interaction, eg. that system responses are conditional on the user's last action and thus inferences can be drawn from an absence of a response in the same way as the noticeable absences of second pair parts of adjacency pairs in conversation, but the machine has inadequate access to the resources that people do, and thus the occasioned breakdowns that occur in expert system use are unavoidable and irrevocable when the machine behaviour is predesigned. The thesis here further recommends the appropriateness of an interactional methodology both for human-computer interaction and computer-mediated communication, and also uncovers some aspects of the taken for granted interactional achievement of dialogue and the emergence of dialogue organisations that is called for if the conception and design of dialogue 'with' machines is to be improved.

6.3 Copresence and Interactivity

The last chapter investigated verbal dialogue in which the main work is the temporal and participatory contextualisation of largely 'given' communication, ie. the contributions are inscriptions worked on and over as explicit and abstract objects. What more complex human interaction dynamics are to be found that may provide an appropriate model for pragmatic dialogue artifact design? This section will emphasize the rich continuum of communication in copresent conduct that could be said to be 'given' and 'given off'. A complex of concurrent and sequential resources are routinely co-ordinated and mangaged in the achievement of shared understanding. The shift in the thesis at this point towards the particulars of full

copresent engagement is not meant as a recommendation for its replication in the design of artifacts. Rather it is to rectify the problematic emphasis placed on focused verbal talk as the most relevant source of insights. Kendon (1988, p. 35) notes that the amount of work on non-speech interactions is little compared to the close analysis of spoken interaction, and the following quotation describes some of the problems this can cause.

> *There are at least three ways in which theories of communicative competence are limited by construing speaking as mainly talking: (1) by underemphasizing nonverbal and paralinguistic behavior at the expense of the verbal; (2) by fostering an analytic separation between speakers and their hearers that obscures the social organisation of the face-to-face interaction within which talk occurs; (3) by focusing research attention on social situations in which talk is the central, 'foregrounded' aspect of the activity, and consequently diverting empirical research and theory development away from considering interactional occasions in which talk is a secondary, 'backgrounded' accompaniment to other action.* (Erickson & Schultz 1982, p. 215)

As a consequence, there is a very real danger, if the interchange model is abandoned, that AI dialogue systems will still have to operate on the basis of evidence that is quite insubstantial for people, ie. as if "listening to each other through a keyhole." (Erickson & Schultz 1982, p. 216) An understanding of both routine human dialogue in copresence and the achievement of dialogue in restricted modalities will help us come to an understanding of the foundational features of mutual intelligibility in dialogue activity. This may help design, and also make us appreciate why some dialogue systems do not work when they look ideal on paper.

The discussion that follows considers some features of dialogues that occur in the full conditions of copresence. The experimental study described in Chapters 4 and 5 used a 'constructive interaction' technique to expose the chronic process of inscription production in a restricted virtual dialogue space. Two coparticipants worked together on constructing an intelligible mutually satisfactory route in and through the dialogue space, and thus were required to organise their collaborative participation in order to achieve this within the dialogue space provided. The diagram in Figure 4-1 in Chapter 4 (Section 4.4.1) illustrates the physical layout. The coparticpants were seated side-by-side and directly in front of them on a table were a number of artifacts: a map, a keyboard and a screen, available as resources which which to achieve success in the task. A number of activities conducted with these artifacts had to be coordinated by the coparticipants either as a mutual focus of attention or in parallel. For example, the map was an essential resource for finding the sense of route or location descriptions. Sometimes coparticipants would mutually orient to the map surface. At other times only one, usually whoever sat in the chair next to the map, worked on the map while the other worked over the keyboard or screen. Also, the screen was an essential resource for monitoring ongoing and past dialogue contributions recoverable from the visual surface. Often, one coparticipant would attend to the construction of a dialogue contribution while the

other would scan the map surface. Thus, not only do the parties in the VDS have to coordinate the relevance of their contributions to each other but also coparticipants must coordinate their own collaboration - the copresent coparticipants had a much richer set of resources available as we shall see. But the setting was quite different from face-to-face conversation because almost all of the time eye contact was absent between the coparticipants, though not noticeably for those involved. Coparticipants sometimes worked in silence, at other times in rapid speech exchange characteristic of conversation, but always their attention was directed to a visual and physical workspace and their bodily actions in and around that space. Their talk is embedded in and can only be understood with reference to the tasks engaged in, and the artifacts they work with. Note that silences in their talk are not necessarily significant as absences of activity. Incipient talk is a term for fitful talk while engaged in other, maybe primary, activities. Lynch (1985) has noted that incipient talk is a characteristic feature of scientists' discourse in the laboratory, and it is certainly true in shared offices.

6.3.1 Introduction to the Analysis

The basis for some of the analyses is derived from the empirical findings of Schegloff (1984), Erickson & Schultz (1982), Goodwin (1981), and Goodwin & Goodwin (1987). Videos of coparticipants working together were collected that recorded essential details of their achievement of communication in the VDS. The videos also display the collaborative work of coparticipants in achieving a task together with a range of physical devices, spatial representations, etc.. The examples following are from one long fully transcribed fragment given in Appendix B in which two coparticipants manage their mutual orientation while engaging in a complex coordination of gesture and talk in a disagreement sequence. At the beginning of Example B3, both J and R are working on constructing a contribution. However, while J types the response, R attends to the other party's simultaneous concurrent construction. The problem is that J is still attending to the construction of an appropriate response when R orients to the other party's response in his talk. J misinterprets R's actions as elaborating in some way upon the construction he is still engaged in, and a disagreement ensues. The misinterpretation is understandable because of the syntactic and semantic congruity of the half-constructed contribution, "Our starting point is", and the reading of the other party's contribution as "is the A1303 in the top left hand corner?" The disagreement claim by J, "no it's not. our starting points down here", displays that J understands by R's actions and agreement claims that a collaborative turn completion by the other party has successfully predicted their statement of the starting point and requires a reply confirming this. For J the gestures by R are interpretable as pointing to the starting point as predicted by the other party. But for R they are locating the referent of the other party's question that

is clearly problematic for J. Finally, J recognises 'what R is talking about' and responds in the VDS with "yes", an answer to the question "is the A1303 in the top left hand corner?".

6.3.2 Preliminaries

Three aspects of face-to-face interaction must be mentioned before the brief analyses presented in the next sections: copresence, vocal and body activity, and intra-turn dynamics. First, copresence has already been discussed in Chapter 3 and in Chapter 5 in terms of the restricted achievements of copresence. Basically, "in the course of their daily activities individuals encounter each other in situated contexts of interaction - interaction with others who are physically co-present." (Giddens 1984, p. 64) Second, early research in conversation analysis concentrated on telephone conversations in order to exclude the complexity of the face-to-face setting. Upon returning to copresent conversation, it has been discovered that speech and non-vocal actions of face-to-face communication are highly integrated. The body in interactional contexts is not irrelevant for the constitution of talk, but

> *talk and nonvocal activities are closely coordinated and oriented to by participants in the production and monitoring of each other's actions. If the participants themselves are rountinely sensitive to the ways in which vocal and nonvocal phenomena are intergrated, the puzzle of how they work is unlikely to be resolved by studying one or other independently of the overall context of the sequence of interaction.* (Atkinson & Heritage 1984, p. 224)

For example, eye gaze and gesture are systematically related to speakerhood and turn construction (Schegloff (1984) and Goodwin (1981)). Third, joint communicative action is accomplished through the participants' continuous engagement in speaking and listening, and thus the interactional dynamics within a turn are equally important as those between turns.

> *By making projections about the future course of an utterance, recipients demonstrate that they are not dealing with it as a monolithic whole, or simply as a static string of symbolic components tied together through syntax, but rather as a process that emerges through time and carries with it an expanding horizon of projective possibilities that are relevant to the actions that recipient might engage in while acting as a hearer to the utterance.* (Goodwin & Goodwin 1987, pp. 24-5)

6.3.3 Embodiment of Action

The study of the VDS showed that the monitorable transience of actions yield performance characteristics that are irrecoverable from the traces of dialogue that are left on display. Participants in the VDS can orient to certain aspects of the production rather than treat each other's contributions as transmitted messages. But, the bodily production of those actions is not monitorable by the other party, ie. there are no directly perceived features of a bodily

production of an action, only the traces or imprints of observable events instigated by an actor. The embodiment of action is, however, available as a resource for coordination in copresence. To illustrate something of the resources available in the embodiment of action a small strip of activity between J and R will be analysed, namely the slice marked A on Example B3 of Appendix B. J completes the initial trajectory of the instruction "Our starting point is" and looks to the screen. Because J is still engaged in constructing a contribution giving the location of the starting point he begins to orient to the map for material to complete the construction. The bodily shift and movement of eye gaze also provide R with clues for what J, who sits beside him, might be up to, eg. looking to the map for a next activity, or for material, or to check a referent description. As J's gaze comes to the middle of the map surface, I would guess at Marley primary school, R gestures to the screen and says "oh". R then moves his arm, starts an utterance and looks towards the map. However, R's pointing gesture reaches acme on the map at the top left corner. Rather than the gesture appearing suddenly or 'out of the blue' the embodied production allows J, who has oriented to another part of the map, to reorient to R's talk by tracking the gesture's course. J's head sweeps round with the trajectory of R's gesture and at its acme J's gaze has arrived at the top left of the map. R displays by his gestural identification of "the top left corner" that he interprets that J is ready for the next activity. In fact, J is trying to complete the in-progress contribution from what is already constructed, viz. "Our starting point is", and finds R's actions as elaborating on that.

The embodied production of actions illustrated here also enables people to skillfully coordinate their everyday non-vocal actions as Goodwin (1981) has shown. From a video of two friends in copresence a skilled achievement of mutual action, namely one lighting the cigarette of the other, is analysed. After a delay, one friend brings the lighter towards the cigaratte only to find that the recipient is attending to someone else. The inattention to the joint work of a lighting of her cigarette by her friend is not disruptive as such because it is routinely accomodated into the ongoing action. The friend with the lighter modifies her action to include an attempt to fix the lighter whereupon the lighter is again brought to the cigarette as a joint accomplishment after attention is resumed by her friend.

In Chapter 5 it was argued that computer models of dialogue have adopted the interchange model of dialogue of which there are three consequences. One consequence is that the model does not allow for self-coordination and repair because messages are constructed in secret and then transmitted. There is no monitoring of the embodied performance of action, only the traces remain in the complete product that the recipient must wait for. The next example, the slice marked B on Example B3 in Appendix B, illustrates some dynamics of self coordination that are missing from computer models of dialogue. The focus is on the

utterance "is the a- a one th one three oh three" by R. Just prior to the utterance, R had still not succeeded in achieving agreement with J after claiming "yeah it is yes" plus a pointing gesture to the screen coordinated with the last "yes". R responds to the uneasy silence, a potential dispreferred disagreement, with turn holding "ah", reaffirms, and then begins a repeat of the line by the other party on the screen. The gesture to the screen is withdrawn and after the cutoff the hung right hand pointing gesture is dissolved as R looks away from the screen to check for evidence on the map surface. The map is a source for identifying the referent, given J's potential disagreement. Thus the course of R's action is contingent on the coordination of his own perceptual and bodily activity. The noun phrase is restarted, along with the transfer of the right hand towards the top of the map, but self-interrupted again as his gaze comes to rest at the top left hand corner. It is postponed for a fraction as R scans the map surface and the gesture reaches completion. The gesture is thus more precisely coordinated with the noun phrase. The reading aloud is not itself evidence for identification, so it has to be coordinated with a demonstration of evidence on the map surface. The evidence now found and displayed, R returns his gaze to the screen, closely followed by J. So, in this example, a participant coordinates several complementary actions, but coordination contingencies lead to the self-repair of the ongoing turn construction. It is important to note that in any multi-modal computer dialogue system the mere fact that complementary actions, solely by the machine or user, may be possible and are monitored means that self-coordination will be problematic because of endogenous troubles. This is not to be avoided, for example by planning ahead, but is an essential constraint and resource for situated action. In addition, monitorable embodied actions are publicly available for displaying orientations to one's own conduct.

6.3.4 Complementarity of Actions

Gesture and speech are not necessarily sequentially produced as can be seen throughout the full fragment in Appendix B. Gesture and speech can occur at the same time in actions by more than one person as well as by one speaker, which is unlike the normative organisation of speech itself even though gesture is associated with speakerhood. This is because of the specific characteristics of the modalities and their use, eg. gesture can be hung or delayed in progress but remain visually latent. Speakers can comment on or talk over gestures and activities by other particpants, eg. they can complement each other or occur simultaneously without relevance. Erickson & Schultz (1982) have argued for a distinction between reciprocity - "the interdependence of actions taken successively across moments in time" (p. 71); and complementarity - "the interdependence of actions taken simultaneously in the same moment" (p. 71). The notion of complementarity of actions will

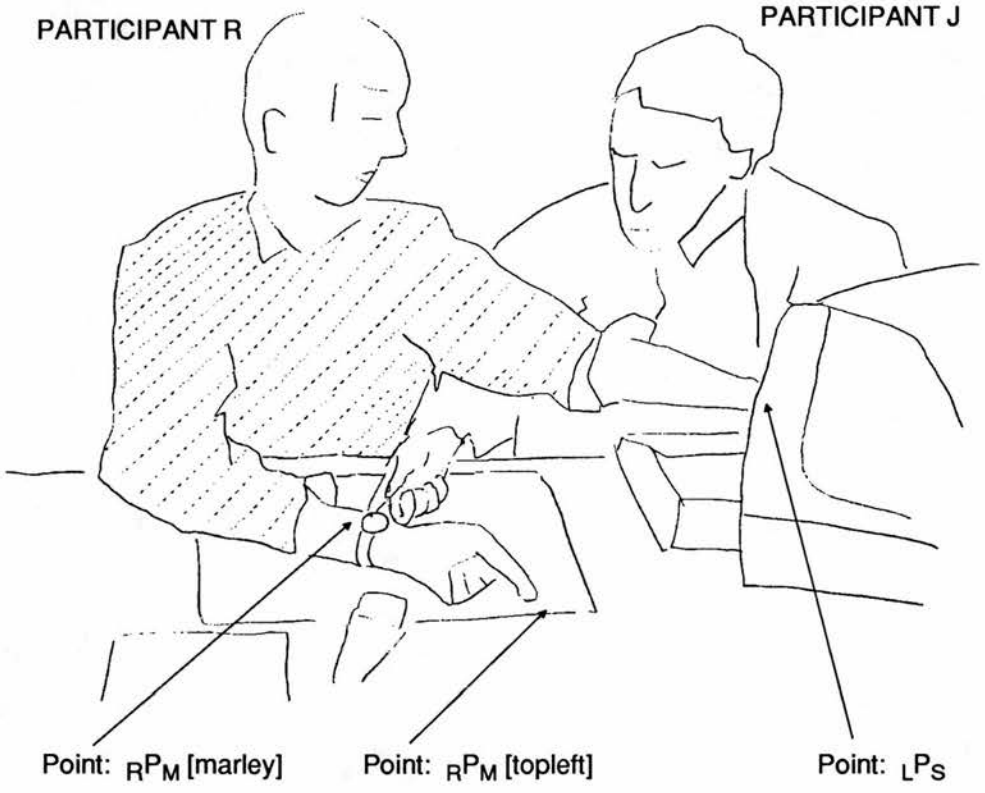
be looked at in this section in contrast to the weaknesses in the interchange model of dialogue, and an example used to illustrate the coordination of complementary actions.

One consequence of the interchange model of dialogue was discussed in the last section, and a fragment was used to illustrate what resources have thus been abandoned by computer models. Here, the other two consequences will be considered. The first is that the model does not allow recipient design in concurrency because messages, if they are constructed in public, are not available for redesign, accomodation nor redirection on the basis of the recipient's behaviour, eg. the recipient must take it or leave it. The second is that the model does not allow for the mutual management of participation because the exchange of messages is strictly controlled, eg. do not take a turn until the speaker explicitly says. The next example, the slice marked C on Example B3 in Appendix B, is concerned with recipient design in cross-speaker concurrency, and gesture overlap and competition. It is a good example of third turn repair and re-explanation because R finds that J's actions display a misinterpretation and so R attempts to repair this in third turn. J begins a turn, after R's prompts to respond affirmatively, with "no it's not". This is coordinated with a shift of gaze to the map and the beginnings of a pointing gesture to display evidence for disagreement on the map surface. However, R simultaneously begins a pointing gesture towards the screen. This could be because of the second noticeable absence of expected agreement or action after R's "Yes. (pause) Yes." and short gesture to the screen. At this point two potentially relevant gestures by different parties and one speech stream are simultaneously produced. The competitive overlap is resolved by R dropping out and monitoring J's action, but still holding the pointing gesture to the screen. In this case, speakerhood is a practical issue, so the occurrence of concurrent gestures is a local problem to be resolved. Immediately after, latching onto J's disagreement, R responds verbally with "NO NONO", and gesturally, thus initiating a repair of the misunderstanding. The sketch in Figure 6-1 illustrates the video image at this point, just before J looks to the screen (marked on Example B3 in Appendix B). The held pointing gesture by R is now read back into the flow of conduct, post-hitch, by reading the screen text into the talk once again. It is withdrawn after the reading is complete. However, during the reading a second complementary display of evidence on the map surface is managed concurrently with the talk. So in this example the coordination of gesture in talk is subject to interactional contingencies. It has to be repaired and accommodated on occasions of concurrent action and recipient monitoring.

The brief analyses in this section are intended to complement the in-depth study described in Chapter 5. Two interesting features of copresent dialogue have been isolated, viz. embodiment and complementarity. Together with the findings of the last chapter they shed light on the complex achievement of dialogic presence and interaction that can inform the

design of intelligible artifacts apparently able to engage in dialogic activities with people. By carefully designing for interactivity a wealth of resources becomes available which people themselves draw upon in the constitution of their actions. For example, the embodiment of action provides a texture of presence and monitorability for routine coordination of interaction

FIGURE 6-1 - A sketch of the video image from [M5:1'48"].



and participation. Also, the complementarity of action furnishes immediate feedback on how the recipient take the action, a resource for mutual adjustment, and a tacit experience that aids memory. These and the other features documented in Chapter 5 not only provide resources but also constraints on interpretation.

6.4 Designing for Interactivity

There are two clear directions in using investigations of human interaction for the design of machines premised on models of dialogue. The first one is to try and reproduce, exploit and adapt features of human conduct for the particular setting and problems encountered, eg. to engineer some interactivity into dialogue artifacts that engage in dialogue 'with' a user. The second is to look at the breadth and particularity of work done and resources available in human collaboration and communication, and creatively design new mediations and resources in order to facilitate dialogue 'through' machines. For example, new types of

synchronous collaboration need facilities for coordinating intelligibility of asynchronous work and movements between asynchronous and synchronous activity.

In the immediate future, cycles of design, engineering and evaluation will clarify possibilities for both directions. This is because of the unintuitive development of dialogue in practice and the contingency of user acceptance. In the long term, understanding and decreasing the asymmetry between human and machine by incorporating foundational aspects of dialogue activity which includes presence, performance, participation, embodiment and complementarity, as well as local adaption and repair, will be a major research development. Much can be gained from studying the micro-dynamics of intra-turn and inter-turn action, eg. the mutual coordination of moments of presence and participation, and this has been the topic of the study here.

6.4.1 Channelling the User

One must design for interactivity in dialogue artifacts but first a note of caution. In order to design for interactivity the user must be channelled or constrained in order to sustain practically the metaphor of dialogue 'with'. The asymmetry between user and computer in terms of access to the resources of action means that user channelling techniques and alternative resources for the system will have to be developed recursively and contingently, with a distinction between particular cases of breakdown and general troubles with a system, ie. some troubles will recur and be undesirable while others will be locally occasioned and can never be designed out. The goal is to avoid or manage for contingency whilst recognising the essential grounding in human practice and the superior interpretative powers of people. It is essential to channel the user given the machine's restricted and asymmetric access to resources of situated action (Suchman 1987). The asymmetry is not only between human and machine interpretative powers, but also in what can be accepted by the machine and what can be produced. eg. the system's responses can be quite complex textual units that are meaningful for a user, but the user can only construct actions using restricted language forms that are suitable for processing.

With regard to user troubles, they can range from unavoidable breakdowns emerging for a user but are not routinely encountered, to repeated troubles that occur among the general user population because of bad or inappropriate design. The first category is implicitly a part of the nature of action itself and artifact use and cannot be avoided nor successfully accommodated by current machine-centred design methods. The second can be pinpointed and maybe resolved in the cycles of evaluation and redesign, eg. by changing the system organisation or explicitly doing channelling work to avoid the problem in situ. Channelling is essential for dialogue systems operating in environments without explicit conventional means

of coordinating participation. For example, speech systems may require explicit means of turn boundary determination or else the system cannot rely on a context-free determination of a speaker's participation. Or the user's experience with the system must be manipulated so as to favour certain responses or interpretations over others. Channelling can be undertaken in different interactional ways with some degree of success. For instance, a designer can try to prevent unwanted behaviours by narrowing the choice of possible user responses, physically constrain the participation of the user, or have the system give standard first pair parts that require a specific type of response. Also, a designer can try to curb the user's overextension of the capabilities of the system by displaying the system's interpretation more explicitly, or designing for a set of restricted resources the same for both the machine and user so the user can 'empathise' and learn what is interactively possible from engaging with the system.

6.4.2 Interactivity in Dialogue 'With' Machines

Up until recently, computational approaches have aimed to produce 'natural language' dialogue systems that allow the user to communicate with the machine using typed language turns through an interchange mechanism. If such a system is viable and useful, the only resources available to the user for constituting the significance of actions is through the sequential organisation of turns and the linguistic resources available. All local interactive work normal to conversation can only be achieved instead through the constraints of an interchange system, eg. participation is fixed, so the management of initiative is weak; a sense of presence and real time monitoring is unavailable, so the local and contingent construction of turns is lost, as is the visibility of self-coordination; recipient design is much more tentative; and adjacency becomes an essential resource which leads to multiple conditional relevances (threads), and more explicit inferences that fail.

The importance of non-vocal and the taken-for-granted achievement of mutual orientation has been acknowledged in theory by Nickerson (1977). Dialogue systems need to be "given the means of conveying to the user the type of information that people sometimes convey to each other in non-verbal ways. In particular, it should be able to do something analogous to saying 'uh-huh' at appropriate times to assure the user that he continues to have its attention." However, a difficult problem for designing interactional artifacts is the openness of action. Boundaries and descriptions cannot be guaranteed by engineering conventional resources for constituting action interpretable by the machine. The advantage in human interaction is that these matters are locally worked on and mutually achieved, and thus, are sensitive to contingencies. Thus there is a need to support dialogue artifacts that can routinely cope with non-mutual presence because mutual monitoring in copresence is a skilled accomplishment

and access by the machine to the routine resources of human copresence is not currently feasible.

Some recent research has begun to design for interactivity. For example, the work of Frohlich & Luff (1989) exploits an understanding of the communicative resources available in human conversation to design new and appropriate interactive artifacts that support in principle a mixed initiative advice-giving dialogue. Of primary interest is their departure from the interchange model of participation. They model sequential features of the dialogue that influence the processing and production of utterances, ie. they attempt to re-instantiate similar organisations to that found in conversation. For example, adjacency pair structures are categorisations motivated by participants' orientations but are underspecified for the purposes of giving a second pair part in a particular circumstance. Because a system must react intelligibly to the user's actions, a proceduralisation has to make explicit just what a current response should be, in practice. That is, the system must respond in such a way as to reproduce instances of the adjacency pair as an orientable structure in the dialogue. Also, Raudaskoski (1989) has explored how repair might be organised in a hypothetical speech telephony dialogue system based on the findings of conversation analysis. An experimental evaluation of a prototype system used a 'wizard of oz' scenario. That is, a person mediated between the system and the user because of the lack of robust speech technology. The mediator rendered the user's speech into a computer readable form and read aloud the computer's response. However, participation was not explicitly designed but emerged within the constraints of the experimental setup. For example, long pauses were common between the spoken turns of the user followed by the mediator because of the time required for typing and waiting for the computer's response. This was not a problem most of the time because the sound of typing was faintly audible on the telephone and thus was a useful resource for coordinating speaking and listening. The typing sound was an embodied feature of the production of actions available for monitoring that provided a texture orientable to by the user.

Let us take a closer look at the participation organisation in Frohlich & Luff (1989). They have attempted to engineer a restricted form of mixed initiative in discrete speaker-controlled moments of opportunity. This is a design strategy because the crucial resources in co-present conduct for coordination and conjoint action are taken as the model for engineering discrete places of opportunity for reciprocal action in the computer dialogue system. The claim in their system is that turn taking is a resource for managing initiative. Interchange dialogues use a simple and strict turn taking system that only gives sequential adjacency as a resource for achieving intersubjectivity. Management of participation and initiative is only possible through explicit linguistic means. In institutional talk, participation is mainly fixed in advance through restricted orders of turn taking that are activity and setting

specific. In conversation, turn taking may be simply described but in actual contexts of use allows an enormous flexibility of participation. In talk, it is feasible for anyone to speak at any time and at the same time, so action must be coordinated and organised so as to achieve sensible participation. However, in Frohlich & Luff (1989) the places are engineered through the constraints of the system and not as an accomplishment of the participants themselves, ie. the recipient can only initiate participation at pre-designed speaker-controlled places, and not, potentially, at any time. In their advice system there are still many restrictions on the participation of both parties that are fixed prior to the interaction.

In engineering systems that are intelligible through the predetermined self selection options, Frohlich & Luff (1989) are relaxing the fixed turn ordering engineered in interchange dialogue systems. The current speaker's opportunity to offer the opportunity for self-selection plus the selection of next speaker through adjacency pairs gives a better resource for coordinating intelligibility than the current speaker selection of next speaker, that must occur in interchange dialogue. It makes explicit two means of turn management derived from studies of human conversation which have interesting dynamics. For example, the synchronicity of monitoring makes time a significant resource for the participants. By creating an environment encouraging synchronous monitoring - reading speed output - the system reported by Frohlich & Luff (1989) can be designed to work on the assumption that, at appropriate points, small chronos time periods are significant in terms of turn transfer, eg. after a given turn-constructional unit, more than two seconds without a response by the user means that the system can continue turn construction. In this case the performance of action provides a resource for constructing a weak dialogue presence that grants inferences based on the temporal flow of the dialogue. This is an example of the gradual recovery in interactive artifact design of the referential and interactive qualities of talk.

The claim by Frohlich & Luff (1989) that the user will find the significance of advice through the sequential organisation of the dialogue as social action appears reasonable, but it needs empirical investigation as recommended here to see if users continue to engage with the machine in that way after a long period of use in a community of other users. Also, it seems wishful thinking that somehow the design of the advice system plus the simple local management of turns can be claimed to guarantee the intelligibility and appropriateness of that 'found' advice. Besides these points, the work of Frohlich & Luff (1989) must be continued in designing better interactive interfaces that move towards artifacts that can react intelligibly in situations of presence. With the advent of multi-modal interfaces the organisation of the participation of the system and user can no longer be considered using the interchange dialogue model, eg. strict exchange of turns, probably with strict modal sequencing. The organisation of participation and the achievement of intelligibility and

collaboration in human communication and interaction must be exploited in the creative design of interactive artifacts.

It was discussed in Section 6.2.2 that interactive technologies should support a user's direct participation and that the impression of continuous engagement is a priority. An analogy with moving image technology was made because they exploit the pecularities of human perception so as to produce continuous flowing images from the projection of static images one after another at a reasonable pace. Similarly the notion of 'interactive flow' in dialogue artifact design should be considered as a goal recommended by the findings of the thesis. That is, systems should support the illusion of continuous participation in which the opportunties for participation by user or machine in the ongoing activity are manifold. Designing for interactivity that supports presence, performance, and embodiment will move towards direct engagement dialogue environments. Consider the opportunity and possibility variation in interactive systems. The normal conception of 'interactive systems' is an interchange reactive environment. By decreasing the grain size of opportunity an 'open' world of interaction can be partially created. An appearance of open participation can be sustained by having small grain moments of opportunity to act and many possibilities for action at each moment. The user perceives not a series of discrete opportunities or at worst a sequence of system controlled and rare opportunities, but a continuum for all practical purposes. The aim would be to support this illusion in the design of dialogue 'with' the machine even though the space of opportunities and possibilities is engineered with discrete materials and the user is constrained by channelling techniques[4]. For example, the design of a telephone dialogue system - a virtual talk space - implicitly allows continuous participation for people, but it is a difficult research problem to design so that the resulting human-machine dialogue will have the appearance of an 'open floor' of participation rather than a strict turn order that must be obeyed or breakdown will occur. The system would have to be responsive to the user's interactive demands such that the user would be able to participate in the intra-turn construction of the dialogue itself.

6.5 Summary

Some consequences of Chapter 5 for computer dialogue artifacts were considered. It was argued that the conception of the artifact must move away from the interchange model but

4 It may be possible to engineer a credible interactive flow by building on the notion of relevance places, ie. that at certain predictable points in a dialogue activity a relevant action by the recipient is expectable. Thus by channelling the user, certain locally predictable points of relevant action will arise that are of an adequate grain size for the user's smooth participation, and yet can be managed by the designed machine based on AI techniques.

not to the adoption of spoken conversation as the model for replication. Instead investigations of the achievement of mutual intelligibility in dialogue in and through contexts of interaction must be conducted, and so the findings of Chapter 5 add to an understanding of a complex interactive practice. Designing for presence and interactivity - a texture of interaction - is essential to avoid over-representation, time-consuming computation, 'out of the moment' planning, and limited evidence and constraint that are characteristic of adoptions of the interchange model. After the success of the empirical methodology, the thesis recommends the application of interaction analysis to computer-mediated communication and human-computer interaction for design and evaluation. A brief study of copresent collaboration over a workspace was described in order to illustrate two more features of dialogic interaction that are missed by the interchange model, viz. embodiment of action and complementarity of actions. Thus the thesis has documented a range of features for consideration if designing for interactivity in dialogue artifacts as well as supporting dialogue 'through' machines. But it is hard to make predictions because of the unintuitive dynamics of dialogue in new contexts and user acceptance and interpretation, so cycles of design and evaluation are required. The asymmetry between the machine and user in terms of access to resources requires channelling techniques to constrain the user's behaviour and interpretations. The disadvantages of the interchange model are multiplicitous, and recent research is beginning to recognise the advantages of designing for interactivity. However it is important that empirical evaluation of user practices over long periods and in a user community be undertaken at many stages of design. It was argued that 'interactive flow' be an ultimate goal if designing for interactive dialogue artifacts that are maximally and locally sensitive to the user's needs.

Chapter 7

CONCLUSIONS

Parry: I don't understand your motives.
Doctor: Why don't you understand my motives?
Parry: Shouldn't I be?
Doctor: Why do you say that just now?
Parry: That's just the way it is.
Doctor: Suppose that's just the way it were not.
Parry: So?
Doctor: I am not sure I understand you fully.
Parry: You harp on the same questions too much.
Doctor: What does the similarity suggest to you?
Parry: I don't get you.
Doctor: Do you wish that you did get I?
Parry: It's about time you showed some interest in my feelings.
Doctor: Oh? I showed some interest in your feelings?

A brief dialogue between two machines from Cerf (1973).

The main aim of the thesis has been to develop an interactional perspective on computer dialogue models, mediation and artifacts. The focus has been on the notions of dialogue adopted in each domain, as well as in interactional studies, and the relevance of findings or theories in one domain to the other domains. Computer models of dialogue were found to be problematic and limiting in a number of ways. From the findings and arguments presented so far, conclusions are drawn in this chapter for the development of more appropriate computer models. In fact, a radical alternative modelling technique is proposed that is more in touch with the interactional perspective drawn from ethnomethodology and hermeneutics. With 'embodied modelling' new forms of dialogue organisation or behaviour itself may be possible that help us understand the nature of dialogue activity and what forms it could take. Thus the thesis is relevant to computer modelling of dialogue in artificial intelligence and computational linguistics as it currently stands and to future approaches. As for empirical matters, the study indicates that much further work is required in order to understand more fully the constraints, resources and dynamics of presence and participation. In addition, the

results of the study as well as the arguments from Chapter 3 are relevant for the design of novel computer dialogue artifacts and mediations, and some possibilities are considered. The use of interaction analysis in the empirical study could easily be applied to other computer-mediated dialogues, and also human-computer interaction as Suchman (1987) has shown. Thus the thesis is also relevant to the field of human-computer interaction. An intricate rendering method was developed for the study of the virtual dialogue space. This suggests that in order to undertake the analysis of complex dialogue in and through machines, technology could help at the different stages of recording, transcription and analysis that are indeed time-consuming and complex. The thesis adds to the development of transcription systems and techniques for the analysis of human conduct. Finally, returning to the theme of dialogue and the machine, much work needs to be done to understand human dialogue and the role the machine can play in supporting or taking part in similar activities.

7.1 Computer Models of Human Dialogue

The interactional perspective was motivated by considering computer models of dialogue. The emphasis in these models on the procedures instantiated in computers to generate behaviours is interesting because of the attempt to account for what happens when we produce and comprehend dialogue, and how a machine could produce and comprehend intelligible dialogue. The concerns here are those models that simulate or try to engage in dialogue itself, and which must consider meaningful action and the nature of practical understanding. However, as argued in Chapter 2, deep theoretical and empirical problems arise when trying to adopt, use and analyse the models and the data supporting them. The internal workings of the machine and the rapid adoption of ad hoc accounts have become too powerful an influence on the conception of what dialogue is and how it should be modelled. One representative theory was critically analysed in Chapter 2 where problems were found with the weak theory of action and dialogue, with the poor empirical methodology, and with the inappropriacy of the model for the conception and design of computer dialogue systems. The remainder of the thesis after the critique is clearly not an attempt at replacing computer models such as that of Grosz & Sidner (1986). Rather, the arguments closely examine assumptions about dialogue and the behaviour of machines, and the study demonstrates the applicability of alternative empirical methods and documents a range of interesting interactional and circumstantial phenomena. Thus the aim is to suggest ways to supplement current models and maybe supplant them in more radical ways.

Let us consider what would happen if we were to follow on in the tradition of computer modelling. The model of Grosz & Sidner (1986), as well as many computer models of dialogue, attempt to bring together non-linguistic notions of purpose and linguistic structure

into a unified processing model. Let us consider the three proposed components of the model. Firstly, the focus on purposeful dialogue activity is to be welcomed, but it is another thing altogether to propose the recognition and coordination of intentions as the basis for dialogue coherence. For instance, it is difficult to show how intentions are recognised in situ because there is little empirical justification that it is a pervasive practice. However, the relation between plans and situated action is still important, as Suchman (1987) has pointed out, and this topic must be on the agenda for the next decade. An intentional vocabulary is useful and thus worth studying in cases where it is brought to bear on finding sense in observed conduct or in dialogue activity, when there are troubles in the talk for example. Secondly, in the processing attentional component the notion of a shared score-board is problematic, because mutual intelligibility is founded on local practices and not on a background of shared knowledge in the form of representations. This aspect of the model (and machine) must be susceptible to supporting presence and the active construction of participation in situ. The attempt to incorporate structure and purpose is interesting but the status of the structures and purposes is unclear. The ascription of segments and intents to the data needs a much better empirical foundation. Thirdly, the linguistic structuring is remarkably poor in density and consequence. A much tighter understanding of the effects of interaction on language structure and vice versa is required. Besides the problems with methodology, the mechanisms that are purported to operate and give rise to segments as real entities in the dialogue are untenable given the evidence presented in the study. In authentic dialogue, multiple structures and interpretations abound, as do the ongoing reconstruction of interpretations and the reclassification of structures. Interaction must be considered a constitutive domain in which interpretations are contingent, jointly constructed and renegotiable.

An accommodating argument could be that one must add the interactional domain as another 'knowledge source' to be coded and represented to aid the interpretation and generation of an utterance. This could be attempted but emergent properties and routine troubles of dialogue can hardly be represented and planned for purposefully, if they are possibilities within the model at all. A better direction would be to explore how one can re-explain interactionally what was appropriated as cognitive or individual. At this point the model begins to split apart: accounts are divided into those that are internally adequate, eg. coded representation, and those that are externally adequate, eg. in performance. Adequacy can no longer be judged internally or from the formal products, such as linguistic strings, but adequacy should be judged from the dynamics of engagement in activities. Thus the suggestions above cannot be modelled in the original sense of internal modelling. Instead they are to supplement the traditional modelling techniques to build novel dialogue artifacts. This is the domain in which 'knowledge rich' computer approaches in artificial intelligence will

achieve some success and the findings of the study here may serve as fodder. However, the models as they stand provide a poor basis for artifact design as they reify the interchange model. Implementing a system may work for limited applications in which analysts agree on the appropriateness of an output as an action or representation, but processing on the basis of intent recognition and planning is time-consuming, computationally intractable and misses the point. So, basic research is required into a system to support an interactional domain, and, in fact, more appropriate models will most likely come from trying to build novel systems and evaluating them in use.

With respect to the empirical methodology, a better analysis of the data examples in Grosz & Sidner (1986) would be desirable, but one would need to study the original set-up from which the examples were derived. The study here has demonstrated that care must be taken with data, and it must be adequately rendered. In the case of the mediated task dialogue, a number of features would need to be documented: the activities of the apprentice, the expert and the human mediator, and the synchrony of the activities, all of which would require audio-visual techniques and machine-logging.

The suggestions made above to patch up the problems and limitations of computer models are biased towards the design of intelligible machines, rather than adequate models of dialogue. Returning to the crucial problem of computer modelling, different architectures, eg. parallel or connectionist, may help but none of the developments of internal machinery can benefit without recognising the importance of the interactional domain. The hope is that computer modelling and a naturalistic empirical methodology may combine the strictures of computational specification with the pedantic empiricism of conversation analysis but it is not clear if the rigours are compatible. For instance, the precision and justification of the turn-taking systematics as a valid set of rules demonstrably oriented to in conversation will not be better clarified by instantiating them as generatory rules. This style of formalisation will only lead to disaster because of a gross misinterpretation of the nature of explanation and rule in conversation analysis (Button 1990). As already pointed out, the specification of procedures to generate behaviours is not where most of the work is to be done; it is in the dynamic interaction between machineries and environments of activity. There has been too much emphasis on explicit representation and control, and there has been a sad performance due to a fundamental misconception of what action is and because of the inadequate provision of resources and constraints. It is argued here that a radical shift in the use of computers for modelling may give rise to a different understanding.

7.2 Alternative Computer Modelling

An alternative computer modelling technique is to radicalise the way computers are used to model communicative activity. To complete the perspective gained by considering situated interaction and interpretation, an alternative modelling paradigm will be put forward and discussed that attempts to emphasize the empirical method, the social and interactional aspects of dialogue, and the situated, emergent qualities missing from current computer models. It is proposed that we reconstruct an alternative computational modelling paradigm that starts with basic philosophical and methodological issues to explore the nature of communicative conduct. The proposed technique is not so concerned with the intelligibility of simulations for a human user, nor with judging the adequacy of a computer model by the integrity of its formal specification of the representations of action and dialogue. What is required is an account of practical and situated interaction and cognition such that the primacy of the individual is surpassed. This could be investigated using machines to see whether embodied machine architectures can engage in some sort of shared activity, what contexts of presence are required, and how features might emerge in different ecologies that are foundational in the development of dialogue activities.

7.2.1 Problematic Notions in Computer Modelling

Several notions that are common currency in computational modelling have come under attack here and in the general literature. Problems begin with the difficult notion of representation and the focus on the individual or cognitive domain of explanation. Plan-based accounts in AI and cognitive science have been criticised by Suchman (1987) and Agre & Chapman (1989). They express concern over the problematic and unrealistic conceptual task of action representation and formalisation required for executing actions. Problems arise with the posited relations between intent, action and events as well as with the apparently simple step of execution. Many routine, practical features of talk are rationalised through explicit mechanisms, others are missed or ignored, and this appears to be the only means whereby properties of the conduct can be said to be modelled or to exist, ie. properties must be explicitly produced and represented by the model or system. To formalise these properties and use them in the form of representations in generatory mechanisms misses the nature of situated cognition and action. Oldman & Drucker (1985) have claimed that there are deep incompatibilities between the generatory models of artificial intelligence and the 'toolkit' models of ethnomethodology. But alternatives are not quick to appear. This is may be partly because of the mind-wrenching task of overcoming the rationalist tradition and our everyday way of thinking about action, and partly because the

interpretative and hermeneutic traditions on which the criticism is based espouse a doctrine that is not easily reconciled with a methodology based on computers.

In contrast to the isolated cognitive processor, which is to engage in dialogue 'with' a person, one recent development in artificial intelligence, namely distributed artificial intelligence (Bond & Gasser 1988), has begun to explore what is possible if there is more than one machine, and thus acknowledges that one machine does not have to process or represent everything. The original idea of the distribution of computation on the basis of resources such as time or information gave rise to the notion of a multiple agent system with communicating agents cooperatively solving a task or problem. Metaphors from the social sciences, eg. human organisational structures, have been borrowed for the design of control mechanisms. One of the first approaches to consider dialogue and the distribution of knowledge between robots was Power (1979). Rather than build computer programs that model dialogue by attempting to engage in dialogue with people, his study simulates the behaviour of people engaged in cooperative problem solving. To understand the difference, consider the quotation introducing this chapter that shows the result of two models that were built to engage in dialogue 'with' a user, in 'dialogue' with each other. In Power (1979) two robot agents have to cooperate in order to open a door by means of verbal communication. Other approaches have been suggested, for example Kiss (1988/86) and Storrs (1988). Unfortunately the metaphor of communication is rather weak, as argued in McIlvenny & Raudaskoski (1990), and the techniques used are subject to the objections made above, but it contains the seeds of the alternative modelling technique proposed next.

7.2.2 Alternative Concepts

> *An ant [hypothesis: man], viewed as a behaving system, is quite simple. The apparent complexity of his behaviour over time is largely a reflection of the complexity of the environment in which he finds himself.* (Simon 1970, p. 24)

Consider the difference in the computational machinery required for a machine or entity that gets on and interacts with its contingent world, and an interactive artifact that is intelligible for a user. Artifacts are designed for human consumption and rely on a pre-understanding by the user - eg. the social context and the user's experience with machines and the particular machine - and on channelling by both instruction and by dialogue design, whereas machines in distributed artificial intelligence are without this resource, and that is why they are not so 'intelligent'. This is the basis for the suggestion made here: the emphasis must be shifted to the requirements - in terms of environment, machinery, and dynamics - for machines to engage in interesting practical activities with other machines.

Important concepts that are raised by this thesis, in the line of other studies in this area, are: embodiment and presence, and interaction and participation. Emergence and dynamics have already been discussed, and they will be mentioned again here. Dynamics is a description or a point of view of patterns of activity, eg. many structures emerge and so cannot be attributed to an individual machinery, and many are drawn upon and reproduced unwittingly in situated action and so cannot be reified as governing elements. By investigating and giving the dynamics of activity prominence, the aim is to go against accounts that put the emphasis on social structure, ie. creating the 'judgmental dope' (Garfinkel 1967), and accounts that emphasize the symbol processing individual, ie. the 'cognitive dope'. (Heritage 1984). They are not to be wholly replaced; the important constraints of social structure and cognitive machinery are recognised. The goal is to add a new dimension. Rather than being reified as the generative mechanism behind the coherence and order of action, as in computer models of dialogue, planning should be seen as emerging, most likely in trouble, in embodied and situated routine activity. Machinery has an architecture, and we should motivate the search for reasonable and simple architectures by always referring to the interaction dynamics that machineries participate in, and by comparing them to human conduct (Agre 1988).

One issue is whether or not the findings about human conduct can be re-embodied or realised rather than simulated in the sense of building embodied machines, that is, made living as opposed to being pinned down and discursively objectified by analysts[1]. Another tricky question is: does this proposed methodology imply that human behaviour is codifiable, formalisable or programmable? We should be able to avoid this because it falls into the trap of talking within the terms of the rationalist tradition that gives sense to the traditional models. Woolgar (1987) considers the arguments against cognitivism and concludes that there is a serious danger "that anticognitivism will merely reproduce the assumption of codifiability that characterizes the cognitivist position" (p. 323). Later, Woolgar notes parenthetically that codified means "formalised, reduced to a series of rules such as instructions or to an algorithm." (p. 325). He seems to argue that any possible explanation using computers must require codification of human conduct. The approach suggested here does not try to explain solely in social or cognitive terms, but seeks to explore the emergence of primitive behaviours from interactions between embodied machines - no claim is made for a reproduction of human social conduct, though insights may be forthcoming. So, the embodied machine need not

1 See Pattee (1988) for a discussion of this question in the context of artificial life research.

necessarily embody a codification of social behaviour or processes in order to participate in activities.

7.2.3 Machines, Interaction and Activities

Evidence for the possibility of interesting computer models of activity can be culled from the studies of behaviours of organisms that we have available. Some investigations of animal behaviour, such as that of apes, suggest that foundational aspects of human interactional conduct could be found in lower forms of life (Haimoff 1988). Thus designing for an ecology of embodied machines, a laboratory world of animated entities could be profitable[2]. The argument runs as follows. Artifacts and embodied machines could be present, and thus potentially dynamically active, in the world and with 'others'. Rather than use 'out of the moment' resources, why not find ways of using the world as a positive place to act in, and ways of interacting with 'others'. The machine can involve gears, programs or whatever, but the big difference from human or animal studies is that the internal sub-personal components can be manipulated to investigate the properties that emerge. This could be a useful testing ground for conversation analytic and ethnomethodological ideas at a basic level, viz. proto-practice and machine-specific 'ethno' methods. Such dynamics as troubles and repair, dense body interaction, and the coordination and distribution of resources can be investigated as emergent properties. By looking at human embodiment, some ideas can be gained. There are time-space constraints on human and machine activity limited by the nature of embodiment and the physical contexts in which activity occurs, eg. there are strict limitations on capabilities of movement and perception; time is a scarce resource; multi-task engagement is bounded; movement in space is also movement in time; and there is singular occupation of physical space (Hagerstrand (1975)).

7.2.4 Dialogue 'Between' Machines?

Recent research has begun to study solitary embodied activity (see Agre (1988)), but the further study that this thesis leads to concerns the topic of communicative action and intersubjectivity in practical collaborative activity. The main technique would be to explore the emergence and development of human-like organisations, especially communicative activities, with more than one machine entity. Preliminary, but narrow, questions about dialogue - linguistic communication - between rational agents have been discussed in a recent workshop reported in Galliers (1988) which suggests that "for any agent, human or

2 See McIlvenny & Raudaskoski (1990) for a discussion of the continuity hypothesis about the nature of practice and the possible status of embodied machines.

machine, operating in the ever changing, and unpredictable real-world, the inherent flexibility and expressiveness of linguistic communication is essential" (p. 11).

There are many problems with such a broad scheme of modelling, but many features are motivated by the work in this thesis. For example, the notion of embodiment raises the difficult question of what is being embodied and how. Is it the observable behaviour that is similar to human behaviour, or is there some claim for 'experiential reality'? Can useful research be done by looking at societies of machines that 'collaborate' and 'communicate'? Or will the design of machinery in interaction with its world and 'others' be a useful testing ground for the notions of practice and social action in ethnomethodology and conversation analysis? This is a tricky and contentious issue because it is not clear that computational machinery is adequate for engaging in practical activity that is human-like.

7.3 The Empirical Study

The empirical study is successful in demonstrating a range of constraints and resources that parties use in achieving and maintaining an intelligible dialogue in the virtual dialogue space (VDS) from materials at hand. The circumstantial and situated features - temporal, social or physical - of the constitution of dialogue are essential for intelligible action and this has been taken for granted or ignored by many computer models of dialogue in both theory and empirical methodology. Also, it has been shown that interaction analysis is an appropriate technique for the study of computer-mediated dialogue and that an audio-visual study was required in this case. A more naturalistic study would have been better but given the difficulty in tracking the bodily use of the modality in separate locations the experimental setting was required. If authentic data had been available, or another CMC had been studied, then the documentation of its use in different social contexts would have been possible. Unfortunately the experimental context can take no account of social context or culture that may affect the formulation and solution of any task set. Indeed it may not even be a problem. There are no natural temporal, work, or motivational contexts, and thus the experiment defines the problem and in a way suggests its solution through the resources provided. Experimental factors are present but are not considered relevant in the analyses. The task setting was probably the cause of the lack of intricate and extended double dialogues observed in conversational usages. The constructive interaction technique no doubt influenced what happened compared to the work of one to one participation in the VDS, but the concern was not really with trying to find out what one person would do but how the dialogue gets constructed in the setting provided. Also, the setting was unfamiliar and thus unstable activities emerged; this was intended but it would have also been interesting to compare with extended natural usage over a period of time.

Capturing, rendering and analysing activities in the VDS was a complicated and time-consuming process. Much time was spent on repeatedly inspecting and transcribing the videos and developing an adequate transcription method and notation system. It is recommended that similar work be undertaken with the help of video technicians and that adequate time be set aside. The videos were quite adequate, but for analysing work on the map surface an overhead camera would be useful. Better mixing of the pictures would have greatly eased the transcription process. Several other modalities were recorded but were not used, except for a brief illustration, in the analyses presented here. This reflects the nature of the method followed where what will be important or relevant in the analyses cannot be predetermined.

The analyses were documented with many examples from the corpus. It would have been a small improvement if the full transcripts or an extended fragment of a minute or two could have been presented . But this would have been difficult because one minute would take up approximately six to ten pages. A more serious point is that there were no examples of extended double dialogues, ie. with more than two swaps of topic, in the corpus. In fast conversational use, extended double dialogues were observed, and if such dialogues could be recorded then the onset hypothesis could be tested more fully. For future work, a look at different tasks and how they affect emergent orders of participation would be interesting. Also, a possible direction would be to examine the same corpus in terms of the situated work of constructing mutual routes for all practical purposes with indefinite representations, eg. the maps, and how the construction is managed through the dialogue interaction. Such a study would add to an understanding of how representations are used to construct other representations in reference to some future course of action.

The techniques of interaction analysis can of course be extended to the investigation of other modalities. Studies could be made of novel computer-supported dialogues and the resulting emergent orders of participation, eg. the telepen. Other orders of parallel participation such as triple dialogues may be possible. It would be relevant to human-computer interaction to understand how people deal with novel, unstable or troublesome modalities in situ and over periods of time in a community of other users. Also, the techniques could be extended to multiple party CMCs. One possibility would be to examine Grosz's original set-up from which the corpus was derived, eg. the mediated dialogues between apprentice and expert. In this case, and in general, the demonstrable orientation to an intentional vocabulary, rather than the analyst's ascription of purpose and intent to the data, could be investigated, eg. in repair and troubles talk.

The study of copresent collaboration over a work surface addressed only some aspects of the interactional features of that conduct. A deeper treatment of the complex but routine

coordination of embodiment and complementarity should be a long term research issue. Further studies need to elaborate on the ways in which embodiment is both a constraint and resource and what perceptual and physical dynamics are involved. The interactional organisation of complementary actions and activities in dialogue requires a great amount of study to reveal how simultaneous and concurrent actions are interleaved with talk and other activities. For example the sign language of the deaf is a rich visual-spatial language with many simultaneous features that may offer interesting insights for novel artifacts in human-computer interaction. Studies can also be made of collaborative embodied activities such as moving a large object together. Much further research is required to understand how people organise their activities, in and through settings, artifacts, and workspace, which may involve distributed or mobile dialogue engagements.

7.4 Application to Dialogue Technology

It has been argued here that current computer models of dialogue provide a poor understanding of dialogue for conceiving and designing dialogue artifacts. Instead, the study conducted in the thesis and the arguments of Chapter 3 have important consequences for dialogue 'with' machines. One must abandon the interchange model and design for routine presence and participation. What computer dialogue systems might benefit or develop from this perspective?

7.4.1 Dialogue 'With' Machines: Intelligible Dialogue Artifacts

The advantages of designing for interactivity have already been discussed in Chapter 6. Recommendations are easy to make; what is required and is not undertaken here is the implementation of interesting, novel computer dialogue systems that break the interchange model, and draw upon the resources and constraints of supporting participation and presence. In design and successive evaluation one needs to consider: when can a user act? How is the 'act' seen to have, or come to have significance in terms of the system's re-action? Equally, how is the system seen to be re-acting or changing, in terms of its design or of the user's prior actions? More study is needed to understand the relations between language use and interpretation in machine environments compared to other forms of language use, eg. from manuals, signposts and help systems, to sophisticated natural language artifacts in artificial intelligence. It is not yet clear in what practical circumstances the dialogue metaphor is oriented to, and thus care must be taken to support the metaphor when it is useful or needed. Also, innovative work will begin to determine how far the machine can be given access to the resources of dialogue and action without the need for external restrictions.

There is always the danger that opening up opportunities and possibilities for action may raise problems of misinterpretation and the overextension of the machine's capabilities.

Applications of the interactional perspective range from speech or keyboard systems in artificial intelligence to multimodal interfaces in human-computer interaction which incorporate gesture, speech, sound, video, and text, among other media, that can be simultaneous. Domains could include telephone advisory dialogue[3] and pedagogic systems. Cycles of design and evaluation and the novel advances of technology make it hard to predict what will happen and what the problems will be. However, for 'intelligent' multimodal systems it would not be sensible to allow only the sequential production of events in just one modality, eg. a referential pointing gesture followed by a spoken utterance. Thus there is an intrinsic problem of self-coordination and overlap. One is immediately confronted with the possibilities of concurrency and sequentiality in achieving system intelligibility through interaction. The features of dialogue interaction documented in Chapter 6, viz. embodiment and complementarity, are extremely important to understand. If visual, audio and spatial, as well as other modalities are being used and juxtaposed, the system must be designed to allow the user to coordinate 'listening' and acting in the dialogue supported by the machine. Some attempts have been made to incorporate artificial pointing gestures into an interface that integrates verbal (typed) descriptions and pointing gestures for referent identification (Schmauks 1987). Schmauks & Reithinger (1988) have also tried to implement a dialogue system that simulates pointing gestures using advanced graphics; thus the machine can apparently point on the virtual screen. Neal et al (1988, p. 819) report the future "development of intelligent interface technology that integrates speech, NL text, graphics, and pointing gestures for human-computer dialogues." Their focus is on "deictic pointing gestures with simultaneous coordinated NL in both user input and system-generated output." (p. 819) This work is primitive and mainly concerned with input and output, but now future work must confront the possibility of complementary modal events in interaction.

One might hope for media independent resources, structures or procedures that can operate on any media to deal with presentation, participation, or troubles. Given the situated and circumstantial nature of intelligible action and the special characteristics of each media, it seems unlikely that this will be the case. Also a difficult question arises: what remains for global management and traditional AI techniques? An initial step is to move away from work on pre-anticipation, control and planning, and focus, whenever possible, on machine

3 One such system is currently undergoing construction as reported in Gilbert et al (1990). It is to be designed and evaluated drawing on the techniques and findings of conversation analysis in the context of a traditional artificial intelligence implementation.

behaviour that is local, routine, improvised, circumstantially saturated, and interactional. It is unlikely though that traditional techniques will be able to create an 'interactive flow', ie. an appearance of being able to act and participate in dialogue when the moment and need arises.

7.4.2 Dialogue 'Through' Machines: Computer-Mediated Communication and Virtual Spaces

As the study conducted here most immediately concerns a specific computer-mediated modality available on some computer systems, the question then arises of whether the study suggests how the modality could be improved or extended. It is difficult to recommend design features as the investigation was to document not how dialogue should be conducted but the possibility and achievement of dialogue in the first place. One could claim that what is required to inhibit the emergence of unusual orders of participation and the troubles engendered is a set of conventions for turn-taking or a physical embodiment of those conventions, eg. according to the interchange model. This requirement is acceptable but it denies the local opportunities for constructing participation; rather, if there is a problem, then internal resources must be provided for. It is true though that education in the use of and standardisation of practice helps, but ultimately, for elaborations to synchronous modalities, the examples of electronic mail and the telephone suggest that simplicity and local control are very important.

In the VDS modality, the permanence characteristics of the medium combined with the lack of resources for mutual monitoring, allowed parties to construct and address those constructions as ongoing contributions to dialogue or as contributions to be read later. Thus a construction could be treated by a recipient as it is transiently performed with emergent features, or produced as a product - in this case, an inscription on the screen. Such a scenario is difficult to imagine in the speech medium because of its peculiarities, ie. speech is a linear, transient event with no visual spatial residue. But a computer system is feasible through which parties could produce speech documents intended for hearing later but which are monitorable by the other party at the time of their production. Thus, parties could act as speakers and hearers as if on the telephone or shift to producing recorded speech messages and playing the other's messages back when ready. However, the temporality of play-back and the lack of an analogy with spatial simultaneity means that the numerous past contributions cannot be perused at a glance.

In any modality that exploits the permanence of a medium, the performance characteristics that are inherent in transient action are not necessarily intrinsically represented in the permanent medium. This problem is faced by transcribers of human action because any

representation of conduct will be faithful to some aspects of the original conduct and completely ignore others. What is important about the flow of action for the participants may not easily (if at all) be made permanent and readily accessible, so as to be drawn upon in retrospect. Can this finding be utilised in building a CMC that exploits permanence and transience as a differential resource? This could be a possibility in the graphic medium. For example, imagine a shared, graphic space on which parties could write, scrawl, sketch or scribble at the same time though they might be spatially separate. Instead of a permanent record of the inscriptions, as on a blackboard or piece of paper, the medium could be adjusted to fade in different ways, the simplest being a continuum from immediate fade to permanent. Also, some degree of mutually accessible embodiment would be required, such as a ghost of the hand doing the marking on the virtual surface. Parties could write or draw in two dimensions and they could control in some local, mutually agreeable way the degree of fade, and thus permanence and transience, of an inscription upon its production. Because of the participant controlled degree of transience, the mutual monitoring of graphic expression with conversational characteristics is a possibility with emergent events such as the pace and rhythm of the mutually monitored graphic stream. Also, because of the participant controlled degree of permanence, characteristics of writing could be explored, eg. parallel activities and the product. Properties such as double dialogues would no doubt emerge from the practices of parties as they contingently organised their participation through the modality.

Another possibility is to give the machine more of a role in supporting the mediation of dialogue between people. Oviatt (1988) considers whether a computer-interpreter mediation between two participants speaking different languages is feasible and what would be required. She argues the case for considering brokering - dialogues between the interpreter and one of the speakers - as a powerful resource for managing the dialogue, instead of the strict passing of machine translations from one speaker to another. An interaction analysis of how this is achieved in authentic human contexts, for example on the telephone, would be interesting. Unlike the problems with the virtual dialogue space, the limitations and troubles engendered by this sort of technology are deep and socially critical. For instance, it is reported in Miike et al (1988) that a Japanese prototype dialogue translator based on the interchange model and mediating the dialogue between Japanese colleagues and English visitors to a trade show resulted in extended repair initiations by both parties that led into never-ending confusions and a general breakdown of 'trust'.

Extensions to the computer generated simulation of the shared graphic space used in the study can be considered with the advent of recent products such as 'artificial realities' that integrate computers, television, audio and body tracking technology. The question is: how will people cope with machine-mediated communication in a human-made

computer-generated world? Such a possibility raises fundamental issues about the requirements and nature of presence and participation. Before these virtual spaces are ever going to be constructed sensibly, research on these basic features of communication is needed and the thesis makes a contribution in that direction[4].

7.5 Interaction Analysis: Methods

The use of interaction analytic methods was recommended for further studies of computer-mediated dialogues, as well as for the evaluation of implemented dialogue artifacts. Such an empirical paradigm is essential lest ad hoc theorising and misconceptions of what dialogue is, what the machine is capable of, and what exactly goes on when a user community uses a dialogue artifact in its daily practice, prevail. One must pay close attention to seemingly insignificant details which an abstract account such as Grosz & Sidner (1986) must ignore in its specification of action sequences and conditions, by using whatever means are appropriate to record the work of engaging in the activity as intelligible activity. The study here had to develop special notation and record conventions to enable the representation of important features of the dialogue in the VDS for the purposes at hand. The technology of video provided the means to record fleeting events in two separate locations, and the nature of the dialogue as typed inscriptions with some properties similar to spoken talk made it reasonable to adapt the methods of conversation analysis. Unfortunately, the transcription method is not generalisable to the study of the use of computer artifacts, ie. something that would be a justifiable record of the work done by a user of a computer interface.

A conclusion from the study, and in keeping with the theme of dialogue and the machine, is that technology can provide tools for doing analysis in future. Computers can be used at each stage of a study: preparation, recording, transcription and analysis. Some researchers are investigating this - MacKay (1989) and Have (1990) - and the complex study undertaken here also motivates further research. There are of course many problems to overcome and, of course, the dangers of too much unhelpful and unnecessary detail are ever-present. The search for accountable details in conduct - what is real for the participants in the practical setting - must remain a guiding principle and cannot be overtaken by technology.

4 An interesting side effect is that automatic records of many aspects of the embodied performance of action in dialogue activity would be obtained from the computer sensing and tracking of a person's real body, eg. eye gaze, arm gestures, and body orientation. Thus it would prove possible to see, not through the eyes of a person, but from the same physical point of view of any participant or entity.

7.6 Dialogue and the Machine

The purpose of the thesis was to show some of the ways in which the notion of dialogue has been used in our machine age, and to show that, in fact, new dialogue activities are possible with computer-mediation. It is important to see that our theoretical notions of what dialogue is are very much shaped by the devices used in the investigation and explanation of that conduct. Moreover, in some cases the devices are to engage in the practice they are to model and thus demonstrate the model's credibility. With artificial intelligence, research on the nature of human dialogue has a new tool in its armoury, viz. the computer . Now dialogue with the computer can perhaps only be understood once it can engage in dialogue about the nature of dialogue itself. However, in order to better understand human dialogue and the role of machines in supporting or engaging in dialogic activities we must take an interactional perspective and examine empirically and critically the computer models, mediations and artifacts that arise.

Appendix A

TRANSCRIPTION NOTATION

The structure of this appendix is borrowed and modified in part from the set developed in conversation analysis, especially by the efforts of Gail Jefferson.

TALK:

The **HELVETICA FONT** is used for talk and gesture transcription.

SEQUENCING SYMBOLS

Some of these can also be used for marking keyboard activity too.

Overlapping utterances -

A:	ooooooo	
B:	[oo	Onset of overlapping utterances. ([)
A:	oooo]oo	Termination of overlap. (])
B:	[oo	
oo//oo		Overlap onset point in overlapped speech stream.

Continuous utterances -

oo=oo		Chaining of utterances (by one or more parties)
or A:	oo=	without an interval.
B:	=oo	

Intervals of no speech or absence of activity -

(0.5)	Time of pause to nearest 0.5 second.
(.)	Micropause of less than 0.2 second.
(pause)	Untimed pause in activity.
(longpause)	Untimed pause of long duration (>2 seconds).

CHARACTERISTICS OF SPEECH DELIVERY

Intonation -

a) Simple impressionistic set:

.	Falling intonation.
,	Continuing intonation.
?	Rising intonation.
!	Exclamatory intonation.

b) Prosody. Occasionally a linguistic prosodic mark appears, eg. a tonetic nucleus or contour pattern, eg. \ marks a distinctive falling contour.

Sound -

Standard orthography with occasional pseudo-lexical impressions.

:	Extension of sound beyond normal duration.
-	Glottal stop, cutoff.
° °	Passage is quieter than surrounding speech.
CAPS	Passage is louder than surrounding speech.
.t	Mouth click, like in "tch tch" or "tut tut".
{ }	Distinctive reading or quotation voice - eg. what is being read or written is spoken out loud.
<	Hurried start.
h	Outbreath.
.h	Inbreath.

TRANSCRIPTION DIFFICULTIES

(word) or (C:)	Uncertain speech or person.
( )	Untranscribable speech or unidentified person.
(Ø)	Uncertain whether speech or sound.

TRANSCRIPTION DESCRIPTIONS

((descriptn))	Stage directions or comments on the recorded data fragment.
« »	Marks the passage which is commented on by the stage directions.

PRESENTATION SYMBOLS

→	Attention is drawn to this utterance.
1 A: ooo	Line numbering of utterances for referencing in the text.
[code]	Reference code for data fragment in the form [Mn: time1-time2], where n is the experiment code, and the fragment is taken from time1 to time2 recorded on the video tape with an electronic clock counter.
...	Omitted material.

SPECIAL CONVENTIONS: GAZE, KEYBOARD AND GESTURE

GAZE

These symbols mark a transition in the transcript record.

Movements are not transcribed, only the moment of gaze onset.

s	Screen.

k	Keyboard.
m	Map.
e	Eye gaze; eye contact when coincident eye gaze by participants occurs.
Ø	Neutral space.
(···)	Potential gaze, but interrupted or incomplete.
- - - -	Marks a complex passage that is not notated.

KEYBOARD DISPLAY

The COURIER FONT is used to indicate keyboard activity on turn-based records.

Standard character set is from an Acorn BBC keyboard.

•	Space.
®	Return or new line.
[ ]	Character(s) not visible on display - either hidden or not recorded on the display.
// //	Characters within the // are deleted in the reverse order to original typing.
↑	Up-Cursor movement key.

GESTURES

Boundaries of gesture -

((	Onset of gesture.
))	Close/retraction of gesture.

Acme of gesture -

p_m	Point at the map.
p_s	Point at the screen.
g_m	Gesture to the map.
g_s	Gesture to the screen.
gø	Gesture in neutral space.
[description]	Describes the place referenced by the pointing gesture, or the movement of the gesture.

L/R subscripts can be pre or postfixed to the above symbols to indicate the left or right hand.

(·)	Incomplete point or gesture.
⇒	Mark map.
↓	Sharp point or jab.
° °	Small movement.

Appendix B

EXAMPLES

The examples numbered B1, B2 and B3 on the next three pages are referenced in the text of the thesis in Chapters 4, 5 and 6.

1-1 KP EXAMPLE B1

M6: A&J

0'10"

g A: k m s k s k s ks k s k s k s| (18.0)

k A: can • you • tell • me • he //eh// the • starting • point • ? ®

k J: i • am •at • Marley • Primary •

1'00"

g A: |m s| (6.0) |m s m s m | (12.0) |s m

k A:

k J: school ® to • the • east • of • Hills • road

2-2 KP EXAMPLE B2

M8: K&G/E&R

2'16"

g K:(m)p_m ↓ ↓ ↓ ↓ ↓)) s

g G:(k) s m k s

s K: a region= i mean this has got regions on.

s G: right. so i'll put region, (oh) hang on we got some more what's this?

k G: R e g

k E: [ ] [ ] [ ] C a v e n d i s h • A v e n

s E: o- oh- avenue was it?

s R: again yeah

g E:(s) k

g R:(s)

2-2 KP EXAMPLE B3

M5: M&D/J&R

1'29"

g M:(m) s m s
D:(k) m ((p_m s)) ((p_s))

s M: mm yeah yeah
D: corner} cause if they've got that (there) ( )

k D: • c o r n e r[ ] ®
J: • i s •

s J: what? (°what°)
R: oh the top left corner oh smart YES. it is. yeah it is. yes. ah, yeah is the a-

g J: s k s m[marley] m[topleft] s m[marley]
R:(m) s ((g_s))((m p_m s)) m((p_s s)) (($L p_s$))))R

A A B

1'39"

g M:
D:

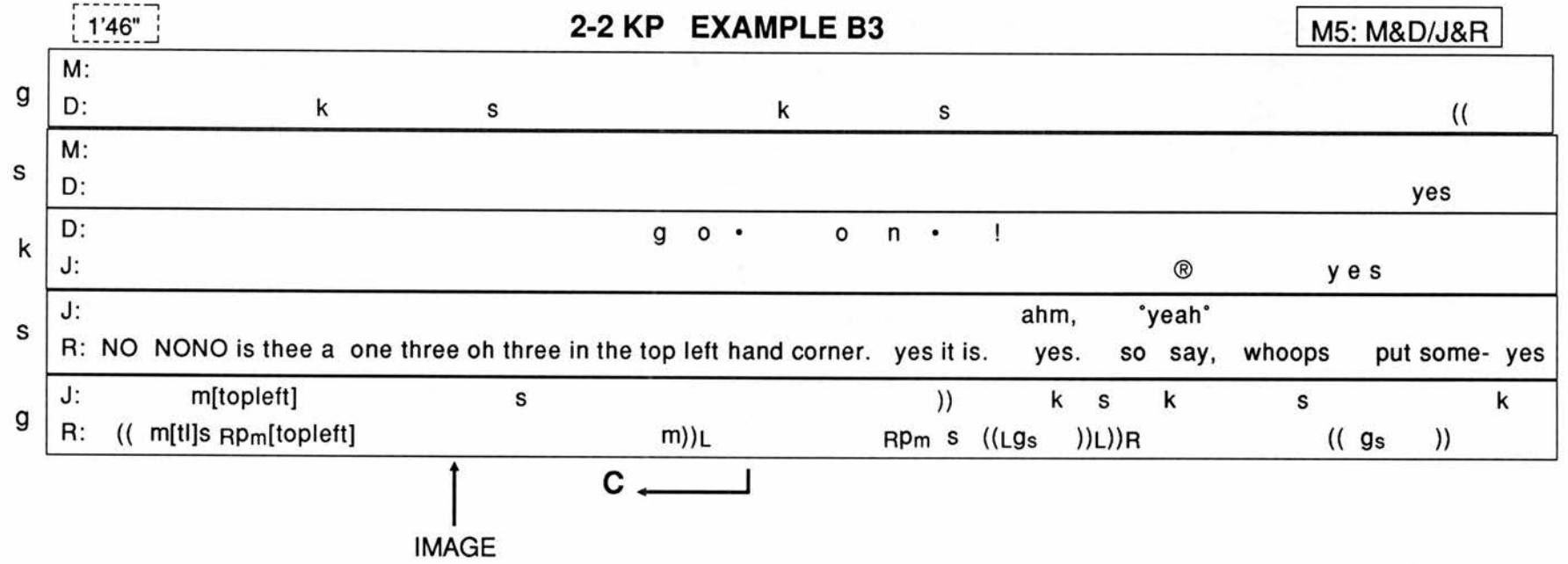
s M: ((laugh)) {our starting point is:} what?
D:

k D:
J:

s J: no it's ˇnot. our starting point's down here.
R: a one th one three oh three on the top left hand corner= Yes. Yes.

g J: m[tl] s m(($R p_m$ [marley]
R: ((m $R p_m$ s)) ((g_s)) (($L p_s$ m[marley]

B C


1'46"
2-2 KP EXAMPLE B3
M5: M&D/J&R
g
M:
D: k s k s ((
s
M:
D: yes
k
D: g o • o n • !
J: ® y e s
s
J: ahm, °yeah°
R: NO NONO is thee a one three oh three in the top left hand corner. yes it is. yes. so say, whoops put some- yes
g
J: m[topleft] s)) k s k s k
R: ((m[tl]s RPm[topleft] m))L RPm s ((Lgs))L))R ((gs))
C
IMAGE

BIBLIOGRAPHY

Agre, P.E. (1988). The Dynamic Structure of Everyday Life. Unpublished PhD Dissertation, MIT AI Technical Report 1085, Cambridge: Mass.

Agre, P.E. & Chapman, D. (1989). *What are Plans For?* Massachusetts Institute of Technology AI Memo 1050a. To be published in Maes, P. (ed), New Architectures for Autonomous Agents, MIT Press: Camb., Mass.

Anderson, A. (1983). Semantic and Social Pragmatic Aspects of Meaning in Task-oriented Dialogue. Unpublished Ph.D. dissertation, University of Glasgow: Scotland.

Anderson, A. & Garrod, S. (1987). *The Dynamics of Referential Meaning in Spontaneous Conversation: Some Preliminary Studies.* In Reilly, R. (ed), Communication Failures in Dialogue and Discourse, Elsevier: North Holland, pp. 161-183.

Appelt, D. (1985). *Planning English Referring Expressions.* In Artificial Intelligence 26, pp. 1-33.

Atkinson, J. & Heritage, J. (eds) (1984). Structures of Social Action: Studies in Conversation Analysis. CUP: Cambridge.

Austin, J. L. (1962). How to do Things with Words. Clarendon Press: Oxford.

Baeker, R. M. & Buxton, W. (eds) (1987). Readings in Human-Computer Interaction. Morgan Kaufman Pub: California.

Bannon, L. (1986). *Computer Mediated Communication.* In Norman & Draper (1986, pp. 399-410).

Bateman, J. (1986). Utterances in Context: Towards a Systemic Theory of the Intersubjective Achievement of Discourse. Unpublished PhD Dissertation. Dept of Artificial Intelligence: University of Edinburgh.

Benson, D. & Hughes, J. (1983). The Perspective of Ethnomethodology. Longman: London.

Birdwhistle, R.L. (1971). Kinesics and Context: Essays on Body Motion Communication. Allen Lane: London.

Black, S.D. et al. (1983). *Real and Non-Real Time Interaction: Unravelling Multiple Threads of Discourse.* In Discourse Processes, 6, pp. 59-75.

Bobrow, D.G., Kaplan, R.M., Kay, M., Norman, D.A., Thompson, H. & Winograd, T. (1977). *GUS, a Frame-driven Dialogue System.* In Artificial Intelligence, 8, pp. 155-73.

Bodker, S. (1989). *A Human Activity Approach to User Interfaces.* In Human-Computer Interaction, 4, pp. 171-195.

Bond, A.H. & Gasser, L. (1988). Readings in Distributed Artificial Intelligence. Morgan Kaufman: California.

Brady, M. & Berwick, R. C., (eds) (1983). Computational Models of Discourse. The MIT Press: Cambridge, Massachusetts.

Brown, G. & Yule, G. (1983). Discourse Analysis. Cambridge University Press: Cambridge.

Brown, J. S., Collins, A. & Duguid, P. (1988). Cognitive Apprenticeship, Situated Cognition and Social Interaction. Institute for Research on Learning, Report No. IRL 88-0008, California.

Burke, J. (1982). An Analysis of Intelligibity in a Practical Activity: The Role and Relationship of Discourse and Context. Unpublished PhD Dissertation, Dept. of Speech Communication, Univ. of Illinois.

Button, G. (1990). *Going Up a Blind Alley: Conflating Conversation Analysis and Computational Modelling.* In Luff et al (eds), Computers and Conversation, Academic Press: London.

Button, G. & Lee, J.R.E. (1987). Talk and Social Organisation. Multilingual Matters: Clevedon.

Carter, K. & Anderson, B. (1989). *Can Video Research Escape the Technology? Some Reflections on the Problems and Possibilities of A.V. Research.* In SIGCHI Bulletin, October, pp. 112-4.

Cawsey, A. (1989). *Explanatory Dialogues.* In Interacting with Computers, 1, 1, pp. 73-92.

Cerf, Vinton (1973). *Parry Encounters the Doctor.* In Datamation, July, pp. 62-64.

Chapanis, A. et al. (1972). *Studies in Interactive Communication: I. The Effects of Four Communication Modes on the Behavior of Teams During Cooperative Problem Solving.* In Human Factors, 14(6), pp 487-509.

Chapanis, A. et al. (1977). *Studies in Interactive Communication: II. The Effects of Four Communication Modes on the Linguistic Performance of Teams During Cooperative Problem Solving.* In Human Factors, 19(2), pp 101-126.

Chapanis, A. (1981). *Interactive Human Communication: Some Lessons Learned from Laboratory Experiments.* In Shackel, B. (ed), Man-Computer Interaction, Sijthoff & Noordhoff: The Netherlands.

Cicourel, A.V. (1973). Cognitive Sociology. Penguin: Harmondsworth.

Cicourel, A.V. (1980). *Three Models of Discourse Analysis: The Role of Social Structure.* In Discourse Processes, 3, pp. 101-132.

Cohen, I.J. (1987). *Structuration Theory and Social Praxis.* In Giddens, A. & Turner, J. (eds), Social Theory Today, Polity Press: Cambridge.

Cohen, P.R. (1984). *The Pragmatics of Referring and the Modality of Communication.* In Computational Linguistics, 10(2), pp. 97-146.

Cohen, P.R. & Levesque, H. (1988). Rational Action as the Basis for Communication. SRI Technical Report 433, Menlo Park; CA.

Cohen, P.R. & Perrault, R.C. (1979). *Elements of a Plan-based Theory of Speech Acts.* In Cognitive Science 3, 3, pp. 177-212.

Cooper, G. (1989). *Ethnomethodology: Its Potential Contribution to HCI.* In Proceedings of the 'People and Computer Interaction Systems' Students Workshop, Open University: Milton Keynes.

COSMOS (1988). Interim Report on the Cosmos Project. Report no. 45.5 EXT, Cosmos coordinators office, Queen Mary college, London.

Coulter, J. (1979). The Social Construction of Mind: Studies in Ethnomethodology and Linguistic Philosophy. Macmillan: London.

Coulter, J. (1989). Mind in Action. Polity Press: Cambridge.

Coulthard, M. (1985/77). An Introduction to Discourse Analysis. Longmans: London.

Daly, J.A., Weber, D., Vangelsti, A., Maxwell, M. & Neel, H. (1989). *Concurrent Cognitions During Conversations.* In Discourse Processes, 12, pp. 227-244.

Dascal, M. (ed) (1985). Dialogue. John Benjamins Pub: Amsterdam.

Dijk, T. A. van (ed) (1985). Handbook of Discourse Analysis 3: Discourse and Dialogue. Academic Press: London.

Douglas, M. (ed). (1971). Understanding Everyday Life. Routledge and Kegan Paul: London.

Drew, P. & Wotton, A. (1988). Erving Goffman: Exploring the Interaction Order. Polity Press: Cambridge.

Dreyfus, H. (1967/72). What Computer's Can't Do. Harper & Row: London.

Dreyfus, H. & Dreyfus, S. (1986). Mind over Machine. Basil Blackwell: Oxford.

Deutsch, B (Grosz) (1974). The Structure of Task Oriented Dialogues. SRI Technical Note No. 90: Menlo Park, California.

Du Bois, John W. (1987). *Meaning Without Intention: Lessons from Divination.* In IPrA Papers in Pragmatics, 1, 2, pp. 80-122.

Duncan, S. (1974). *Some Signals and Rules for Taking Speaker Turns in Conversation.* In Weitz, S. (ed), Nonverbal communication, OUP: Oxford, pp. 298-311.

Edwards, J. L. & Mason, J. A. (1988). *Evaluating the Intelligence in Dialogue Systems.* In International Journal of Man-Machine Studies, 28, pp. 139-173.

Erickson, F. & Schultz, J. (1982). The Counselor as Gatekeeper. Academic Press: London.

Ericsson, K.A. & Simon, H.A. (1984). Protocol Analysis: Verbal Reports as Data. MIT Press: Cambridge, Mass.

Fox, B.A. (1987). *Interactional Reconstruction in Real-Time Language Processing.* In Cognitive Science, 11, pp. 365-387.

Fraser, H.B. (1990). The Subject of Speech Perception: An Analysis of the Philosophical Foundations of the Information-Processing Model. Unpublished PhD dissertation, Department of Linguistics, University of Edinburgh.

Frohlich, D.M. & Luff, P. (1989). *Conversational Resources for Situated Action.* In Proceedings of ACM Computer and Human Interaction Conference (CHI), Austin, Texas.

Galliers, J. (1988). *Report on a discussion session about Dialogue.* In Proceedings of the ALVEY Workshop on Multiple Agent Systems. Redhill, April, pp. 11-12.

Garfinkel, H. (1967/84). Studies in Ethnomethodology. Polity Press: Cambridge.

Garfinkel, H. (1986). Ethnomethodological Studies of Work. RKP: London.

Giddens, A. (1979). Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis. Macmillan: London.

Giddens, A. (1984). The Constitution of Society: Outline of the Theory of Structuration. Polity Press: Cambridge.

Giddens, A. (1987). Social Theory and Modern Sociology. Stanford University Press: Stanford, California.

Gilbert, G.N. & Heath, C., (eds) (1985). Social Action and Artificial Intelligence. Gower: Brighton.

Gilbert, N., Wooffitt, R. & Fraser, N. (1990). *Organising Computer Talk.* In Luff et al (eds), Computers and Conversation, Academic Press: London.

Goffman, E. (1963). Behaviour in Public Places. Free Press: New York.

Goffman, Erving. (1981). *Replies and Responses.* In Goffman, E., Forms of Talk, Basil Blackwell: Oxford.

Goodwin, C. (1981). Conversational Organization: Interaction Between Speakers and Hearers. Academic Press: New York.

Goodwin, C. & Goodwin, M. (1987). *Concurrent Operations on Talk: Notes on the Interactive Organisation of Assessments.* In IPRA Papers in Pragmatics 1, 1, pp. 1-54.

Goranzon, Bo, Florin, M. & Sallstrom, P. (1988). *The Concept of Dialogue.* In AI & Society, 2, pp. 279-286.

Grice, P. (1957). *Meaning.* In Philosophical Review, 67.

Grice, P. (1975). *Logic and Conversation.* In Cole, P. & Morgan, J. (eds), Speech Acts. Syntax and Semantics 3, Academic Press: New York, pp. 41-58.

Grosz, B. (1977). The Representation and Use of Focus in Dialogue Understanding. SRI Technical Report No. 151: Menlo Park, California.

Grosz, B. & Sidner, C. (1986). *Attentions, Intentions, and the Structure of Discourse.* In Computational Linguistics Journal 12, No. 3, pp. 175-204.

Grosz, B. & Sidner, C. (1987). *Plans for Discourse.* Linguistic Institute: AI Approaches to Discourse, L1075 Supplement. Stanford University, Jun-Aug 1987. To be published in Cohen, P.R., Morgan, J. & Pollack, M.E. (eds), Intentions in Communication, Bradford Books: Cambridge, Mass.

Grosz, B, Spark-Jones, K. & Webber, B.L. (eds) (1986). Readings in Natural Language Processing. Morgan Kaufman: California.

Hagerstrand, T. (1975). *Space, Time and Human Conditions.* In Karlqvist, A. (ed), Dynamic Allocation of Urban Space, Saxon House: Farnborough.

Hahn, W. van (1986). *Pragmatic Considerations in Man-machine Discourse.* In Proceedings of COLING (Computational Linguistics conference), Budapest, pp. 520-526.

Haimoff, E.H. (1988). *The Organisation of Repair in the Songs of Gibbons.* In Semiotica, 68-1/2, pp. 89-120.

Hall, H. (1987). *Phenomenology.* In Shapiro, S.C. (main ed.), Encyclopedia of Artificial Intelligence, John Wiley: Chichester, pp. 730-36.

Harris, R. (1987). The Language Machine. Duckworth: London.

Harris, Z. (1952). *Discourse Analysis.* In Language 28, pp. 1-30.

Hayes, P. (1983). *Introduction.* In International Journal of Man-machine Studies, 19, pp. 229-230.

Hayes, P. & Reddy, D. (1983). *Steps Towards Graceful Interaction in Spoken and Written Man-machine Communication.* In International Journal of Man-machine Studies, 19, pp. 231-284.

Have, P. ten (1990). *Text Analysis Programs and Routines: The Challenge of Computer Analysis.* Paper presented in the XIIth World Congress of Sociology conference, Madrid, July.

Heath, C. (1986). Body Movement and Speech in Medical Interaction. CUP: Cambridge.

Heidegger, M. (1967/1927). Being and Time. OUP: Oxford.

Heritage, J. (1984). Garfinkel and Ethnomethodology. Polity Press: Cambridge.

Heritage, J. (1984b). *A Change of State Token and Aspects of its Sequential Deployment.* In Atkinson & Heritage (1984, pp. 299-345).

Heritage, J. (1987). *Ethnomethodology.* In Giddens, A. & Turner, J.H. (eds), Social Theory Today, Polity Press: Cambridge.

Heritage, J. (1989). *Current Developments in Conversation Analysis.* In Roger & Bull (1989, pp. 21-47).

Heritage, J. & Atkinson, J. (1984). *Introduction.* In Atkinson & Heritage (1984, pp. 1-15).

Hirschberg, J. (1990). *Accent and Discourse Context: Assigning Pitch Accent in synthetic Speech.* In Proceedings of the AAAI conference, August, pp. 952-957.

Jefferson, G. (1972). *Side Sequences.* In Sudnow, D. (ed), Studies in Social Interaction. The Free Press: New York.

Jefferson, G. (1983). Two Explorations of the Organisation of Overlapping Talk in Conversation. Tilburg University Papers in Language and Literature, No. 28. Tilburg: Netherlands.

Jefferson, G. (1989). *Preliminary Notes on a Possible Metric which Provides for a 'Standard Maximum' Silence of Approximately One Second in Conversation.* In Roger & Bull (1989, pp. 166-196).

Joshi, A., Webber, B. L. & Sag, I. A., (eds) (1981). Elements of Discourse Understanding. Cambridge University Press: Cambridge.

Kammersgaard, J. (1985). Four Different Perspectives on Human-Computer Interaction. DAIMI PB - 203, Computer Science Dept., University of Aarhus: Denmark.

Katz, J.J. (1966). The Philosophy of Language. New York Press: N.Y.

Kendon, A. (1988). *Goffman's Approach to Face-to-Face Interaction.* In Drew, P. & Wootton, A. (eds) (1988). Erving Goffman: Exploring the Interaction Order. Polity Press: Cambridge.

Kennedy, A. et al. (1988). *Dialogue with Machines.* In Cognition 30, 1, pp. 37-72.

Kiesler, S., Zubrow, D. & Moses, A.M. (1985). *Affect in Computer-Mediated Communication: An Experiment in Synchronous Terminal-to-Terminal Discussion.* In Human-Computer Interaction, 1, pp. 77-104.

Kiss, G. (1988/86). High Level Dialogue in Man-Machine Interaction. A survey report commissioned by the ALVEY Directorate. Reprinted by Human Cognition Research Lab, Open Unversity: Milton Keynes.

Kobsa, A. & Wahlster, W. (eds) (1988a). *Special Issue on User Modelling* in Computational Linguistics 14, 3.

Kobsa, A. & Wahlster, W. (eds) (1988b). User Models in Dialog Systems. Springer-Verlag: New York.

Labov, W. & Fanshel, D. (1977). Therapeutic Discourse. Academic Press: New York.

Laurel, B.K. (1986). *Interface as Mimesis.* In Norman & Draper (1986, pp. 67-85).

Lave, J. (1988). Cognition in Practice. CUP: Cambridge.

Lee, J. (1987). *Prologue: Talking Organisation.* In Button & Lee (1987, pp. 9-53).

Levin, J.A. & Moore, J.A. (1977). *Dialogue Games: Metacommunication Structures for Natural Language Interaction.* In Cognitive Science, 1, pp. 395-420.

Levinson, S. (1981a). *Some Pre-observations on the Modelling of Dialogue.* In Discourse Processes 4, pp. 93-116.

Levinson, S. (1981b). *The Essential Inadequacies of Speech Act Models of Dialogue.* In Parret, et al. (eds), Possibilities and Limitations of Pragmatics, John Benjamins: Amsterdam, pp. 473-492.

Levinson, S. C. (1983). Pragmatics. Cambridge University Press: Cambridge.

Levinson, S & Owen, M. (1981). Conversational Topics and Activities. Final Report of Project on Topic Organisation in Conversation, Jan 1980 - Sept 1981 for SSRC. Mimeo. University of Cambridge.

Licklider, J.C.R. (1960). *Man-Computer Symbiosis.* In IRE Transactions of Human Factors in Electronics HFE-1(1), March, pp. 4-11.

Livingston, E. (1987). Making Sense of Ethnomethodology. Routledge & Kegan Paul: London.

Lynch, M. (1985). Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory. RKP: London.

Lynch, M., Livingston, E. & Garfinkel, H. (1983). *Temporal Order in Laboratory Work.* In Knorr-Cetina, K. D. (ed), Science Observed, Sage: London.

MacKay, W.E. (1989). *An Experiemntal Video Annotator for Symbolic Analysis of Video Data.* In Communications of the ACM 32, 7, pp. 802-810.

McIlvenny, P. (1985). Man-Machine Communication in Dialogues Concerned with Advice-Giving. Unpublished MSc. dissertation, Department of Computer Science, Heriot Watt University, Edinburgh.

McIlvenny, P. (1990). *Communicative Action and Computers: Re-embodying Conversation Analysis?* In Luff, P., Gilbert, N. & Frohlich, D. (eds), Computers and Conversation, Academic Press: London.

McIlvenny, P. & Raudaskoski, P. (1990). *The Practising Machine or the Machine in Human Practice?* In Proceedings of the Workshop on Complex systems, Ethnomethodology and Interaction Analysis, AAAI 90, Boston, July, pp. 31-39.

McTear, M. (1985). *Breakdown and Repair in NaturallyOccurring Conversation and Human-Computer Dialogue.* In Gilbert & Heath (1985, pp. 104-123).

McTear, M. (1987). The Articulate Computer. Basil Blackwell: Oxford.

Mallery, J. C., Hurwitz, R. & Duffy, G. (1987). *Hermeneutics.* In Shapiro, S.C. (main ed), The Encyclopedia of Artificial Intelligence, John Wiley: New York, pp. 362-376.

Martin, J. (1973). Design of Man-Computer Dialogues. Prentice Hall.

Maturana, H.R. & Varela, F.J (1980). Autopoiesis and Cognition: The Realisation of the Living. D.Reidel Pub: London.

Miike, Seji et al (1988). *Experiences with an Online Translating Dialogue System.* In Proceedings of Association for Computational Linguistics Conference, June, pp. 155-162.

Miller, G.A. (1963). *Speaking in General.* Review in Contemporary Psychology, 8, pp. 417-18.

Miyake, N. (1986). *Constructive Interaction and the Iterative Process of Understanding.* In Cognitive Science, 10, pp. 151-177.

Morrison, K. L. (1981). *Some Properties of 'Telling-Order Designs' in Didactic Inquiry.* In Philosophy of Social Sciences, 11, pp. 245-262.

Mulkay, M. (1986). *Conversations and Texts.* In Human Studies, 9, pp. 303-321.

Neal, J. G. et al. (1988). *Multi-Modal References in Human-Computer Dialogue.* In Proceedings of 7th National Conference on Artificial Intelligence, AAAI, Minnesota, pp. 819-823.

Nickerson, R. (1977). *On Conversational Interaction with Computers.* In User-oriented Design of Interactive Graphics Systems, pp. 101-113, ACM: New York. Reprinted in Baecker & Buxton (1987).

Nickerson, R. (1981). *Some Characteristics of Conversations.* In Shackel, B. (ed), Man-Computer Interaction, Sijthoff & Noordhoff: The Netherlands.

Norman, D.A. (1988). The Psychology of Everyday Things. Basic Books.

Norman, D.A. & Draper, S. (eds) (1986). User Centred System Design. Lawrence Erlbaum: London.

Oldman, D. & Drucker (1985). *The Non-Reducability of Ethnomethods: Can People and Computers Form a Society?* In Gilbert & Heath (1985, pp. 144-159).

O'Malley, C., Draper, S.W. & Riley, M.S. (1985). *Constructive Interaction: A Method for Studying Human-Computer Interaction.* In Shackel, B. (ed), INTERACT '84, Elsevier Science Publishers: North Holland.

Oviatt, S.L. (1988). Management of Miscommunications: Toward a System for Automatic Telephone Interpretation of Japanese-English Dialogues. SRI Technical Note 438, Menlo park: CA.

Oviatt, S.L. & Cohen, P.R. (1989). *The Effects of Interaction on Spoken Language.* In Proceedings of Association for Computational Linguistics Conference, June, pp. 126-134.

Pattee, H.H. (1988). *Simulations, Realisations, and Theories of Life.* In Langton, C. (ed), Artificial Life, Addison Wesley.

Perec, G. (1987/70). Life: A User's Manual. Collins Harvill: London.

Polyani, L. (1983). *The Syntax of Discourse.* In Text, 3, pp. 261-270.

Power, R. (1979). *The Organisation of Purposeful Dialogues.* In Linguistics, Vol. 17, pp. 107-152.

Power, R. & Dal Martello, M.F. (1985). *Methods of Investigating Conversation.* Review article in Semiotica 53, 1/3, pp 237-257.

Psathas, G. (1979). *Some Organisational Features of Direction Maps.* In Psathas (ed), Everyday Language: Studies in Ethnomethodology, Irvington Pub.: New York.

Psathas, G. (1986a). *The Organisation of Directions in Interaction.* In Word, 37(1-2), pp. 83-92.

Psathas, G. (1986b). *Some Sequential Structures in Direction-Giving.* In Human Studies, 9(2-3), pp. 231-146.

Psathas, G. & Kozloff, M. (1974). *The Structure of Directions.* In Semiotica, 17, pp. 111-130.

Raudaskoski, P. (1989). Repair Work in Human-Computer Interaction. Unpublished M.Sc. Dissertation, Department of Artificial Intelligence, University of Edinburgh.

Reddy, (1979). *The Conduit Metaphor - A Case of Frame Conflict in our Language about Language.* In Ortony, A. (ed), Metaphor and Thought, CUP: Cambridge.

Reichman-Adar, R. (1984). *Extended Person-Machine Interface.* In Artificial Intelligence, 22, pp. 157-218.

Reichman, R. (1985). Getting Computers to Talk like You and Me. Bradford Book: London.

Reichman, R. (1986). *Communication Paradigms for a Window System.* In Norman & Draper (1986, pp. 285-313).

Ricoeur, P. (1981). Hermeneutics and the Human Sciences. (transl. Thompson, J.B.) CUP: Cambridge.

Roger, D. & Bull, P. (eds) (1989). Conversation. Multi-lingual Matters: Clevedon.

Rumelhart, D.E. (1975). *Notes on a Schema for Stories.* In Bobrow, D. & Collins, A. (eds), Representation and Understanding, Academic Press: New York.

Sacks, H. (1984). *Notes on Methodology.* (ed. Gail Jefferson) In Atkinson & Heritage (1984, pp. 21-27).

Sacks, H. (1989). *Special Issue* edited by Gail Jefferson on Harvey Sacks in Human Studies, 12.

Sacks, H., Schegloff, E. & Jefferson, G. (1978/74). *A Simplest Systematics for the Organisation of Turn Taking for Conversation.* In Schenkein (ed), Studies in the Organisation of Conversational Interaction I, Academic Press: New York, pp. 7-55.

Schegloff, E. (1979). *The Relevance of Repair to Syntax for Conversation.* In Givon, T. (ed), Syntax and Semantics 12: Discourse and Syntax, Academic Press: London.

Schegloff, E. (1980). *What Type of Interaction Is It to Be?* In Proceedings of Association of Computational Linguistics 18th Annual Meeting and Parasession on Topics in Interactive Discourse, p. 81.

Schegloff, E. (1982). *Discourse as an Interactional Achievement.* In Tannen, D. (ed), Analysing Discourse, Text and Talk. Georgetown University Press: Washington D. C.

Schegloff, E. (1984). *On Some Gestures' Relation to Talk.* In Atkinson & Heritage (1984, pp. 266-296).

Schegloff, E. (1988). *Goffman and the Analysis of Conversation.* In Drew & Wootton (1988, pp. 89-135).

Schegloff, E.A. & Sacks, H. (1973). *Opening Up Closings.* In Semiotica, 7, pp. 289-327.

Schiffrin, D. (1987). Discourse Markers. CUP: Cambridge.

Schmauks, D. (1987). *Natural and Simulated Pointing.* In Proceedings of European Association for Computational Linguistics Conference, pp. 179-185.

Schmauks, D. & Reithenger, N. (1988). *Generating Multi-Modal Output - Conditions, Advantages and Problems.* In Proceedings of Computational Linguistics Conference, pp. 584-588.

Searle, J.R. (1969). Speech Acts. CUP: Cambridge.

Sharrock, W. & Anderson, B. (1986). The Ethnomethodologists. Ellis Horwood: Chichester.

Sharrock, W. & Anderson, B. (1987). *Epilogue: The Definition of Alternatives: Some Sources of Confusion in Interdisciplinary Discussion.* In Button & Lee (1987, pp. 290-321).

Simon, H.A. (1970). The Sciences of the Artificial. MIT Press: Cambridge, Mass.

Spark-Jones, K. & Wilks, Y. (eds) (1983). Automatic Natural Language Parsing. Ellis Horwood: Chichester.

Storrs, G. (1988). *Multiple Agent Model of Human-Computer Interaction.* In Proceedings of the ALVEY Workshop on Multiple Agent Systems. Redhill, April.

Stubbs, M. (1983). Discourse Analysis: The Sociolinguistic Analysis of Natural Language. Basil Blackwell: Oxford.

Suchman, L. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. CUP: Cambridge.

Suchman, L. (1988). *Representing Practice in Cognitive Science.* In Human Studies, 11, pp. 305-325.

Sudnow, D. (1978). Ways of the Hand. Harvard University Press: Cambridge, Mass.

Turing, A.M. (1950). *Computing Machinery and Intelligence.* In Mind 59, 256, pp. 433-61.

Walker, M.A. (1989). *Evaluating Discourse Processing Algorithms.* In Proceedings of the 27th Annual Meeting of Association of Computational Linguistics, pp. 251-260.

Waterworth, J.A. & Talbot, M. (1987). Speech and Language-based Interaction with Machines: Towards the Conversational Computer. Ellis Horwood: Chichester.

Weizenbaum, J. (1966). *ELIZA: A Computer for the Study of Natural Language Communication Between Man and Machine.* In Communications of the ACM 9, 1, pp. 36-45.

Wilson, J. (1989). On the Boundaries of Conversation. Pergamon Press: Oxford.

Wilson, J.P., Wiemann, J.M. & Zimmerman, D. H. (1984). *Models of Turn Taking in Conversational Interaction.* In Journal of Language and Social Psychology, 3, pp. 159-183.

Wilson, T.P. (1983). Qualitative "versus" Quantitative Methods in Social Research. Mimeo, Department of Sociology, UCSB, California.

Winograd, T. (1981). *What Does It Mean to Understand Language?* In Norman, D.A. (ed), Perspectives on Cognitve Science, Ablex: Norwood, NJ.

Winograd, T. (1972). Understanding Natural Language. Academic Press: New York.

Winograd, T. (1983). Language as a Cognitive Process: I. Addison Wesley: Reading, Mass.

Winograd, T. & Flores, F. (1986). Understanding Computers and Cognition. Addison Wesley: Reading, Mass.

Winograd, T. (1987). Thinking Machines: Can There Be? Are We? Report No. STAN-CS-87-1161, Dept. of Computer Science: Stanford University.

Woolgar, S. (1987). *Reconstructing Man and Machine: A Note on Sociological Critiques of Cognitivism.* In Bijker, W. E. et al (eds), The Social Construction of Technological Systems, MIT Press: Cambridge, Mass.

Woolgar, S. (1989). *Why not a Sciology of Machines? An Evaluation of Prospects for an Association between Sociology and Artificial Intelligence.* In Murray, L.A. & Richardson, J. (eds), Intelligent Systems in a Human Context, OUP: Oxford.

Wynn, E. (1979). Office Conversation as an Information Medium. Unpublished PhD thesis, Universtiy of California, Berkeley.

Wynn, E. (1980). *What Discourse Features Aren't Needed in Online Dialogue.* In Proceedings of Association of Computational Linguistics 18th Annual Meeting and Parasession on Topics in Interactive Discourse, pp. 87-90.